

CRANFIELD UNIVERSITY

Neal Andrew Barton

Proactively managing drinking water distribution networks: A data-driven, statistical modelling approach to predict the risk of pipe failure

School of Water, Energy and Environment (SWEE)

Ph.D.

Academic Year: 2021 - 2022

Supervisor: Stephen Hallett
Associate Supervisor: Simon Jude
February 2022

CRANFIELD UNIVERSITY

School of Water, Energy and Environment (SWEE)

Ph.D.

Academic Year 2021 - 2022

Neal Andrew Barton

Proactively managing drinking water distribution networks: A data-driven, statistical modelling approach to predict the risk of pipe failure

Supervisor: Stephen Hallett
Associate Supervisor: Simon Jude
February 2022

This thesis is submitted in partial fulfilment of the requirements for the degree of Ph.D.

© Cranfield University 2022. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

ABSTRACT

Water distribution networks are critical infrastructures, providing clean water to millions of people. 3 billion litres of water are lost through pipe failure every day in the UK, impacting serviceability. Statistical pipe failure models can reduce pipe failures by providing valuable insights to enhance decision-making and promote proactive management. This research aims to understand the complexity of pipe failures in water distribution networks and develop a methodology for a reliable pipe failure model that identifies the risk of failure. Through an embedded case study and data-driven approach, several Objectives have been undertaken that comprise the body of research delivered through several research papers.

This study offers several contributions to the immediate field of pipe failure research. Firstly, the findings investigate new factors that form the various modes and mechanisms of pipe failure, using alternative methods not commonly used in pipe failure research are used, including Generalized Additive Model and Dijkstra's algorithm, and using data from a large UK water distribution network. Secondly, the research develops a suitable methodology for predicting annual pipe failures using an advanced machine learning method; a methodology that is easily transferrable. Thirdly, the research provides a useful means of predicting the risk of failure and visualising the results. Fourthly, the research investigates the challenges of pipe failure models using a semi-structured interview approach to review current practice. Finally, the research contributes by exploring several different data-driven methods and an embedded case study design to contribute to the broader context of pipe failure modelling.

The approach presented in this research provides a methodological framework to enhance decision-making for asset management of pipes in clean water networks. Furthermore, it highlights the main limitations, particularly data quality and quantity, data-pre-processing, and model development, highlighting areas for future progress.

ACKNOWLEDGEMENTS

Completing this Doctoral Research was only possible due to the support and guidance of numerous people. I will begin by thanking my supervisory team, Professor Stephen Hallett, and Dr Simon Jude, for their encouragement, support, and advice and for critically reviewing all my work throughout. I would like to thank Dr Timothy Farewell, who played a central role in guiding this research, to Timothy Acland for kindly sharing his extensive knowledge on water distribution networks, and to Dr Trung Hieu Tran for his help with the modelling. I also wish to acknowledge the Natural Environmental Research Council, NERC, and the DREAM CDT. for supporting me. I would like to thank Cranfield University and the School of Water, Energy and Environment staff. I would like to thank the water industry professionals who took the time to share information and accommodate requests and those participants who kindly agreed to discuss pipe failure modelling to help advance this research.

I would like to thank my family, who have weathered the emotional highs and lows inherent in the Doctoral process. I am especially thankful to my brother Marc for being a valuable sounding board, given his experience, and for always supporting me. I am eternally grateful to Kirstin for her relentless patience and emotional support, which has kept me going to the end, and for giving me the space and time to dedicate to this research. Finally, I would like to thank my children, Sebastian, Rafe, and Mila, for bringing me joy and giving me the motivation and purpose to finish this research.

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS.....	ii
LIST OF FIGURES.....	vii
LIST OF TABLES	xi
LIST OF EQUATIONS.....	xiii
LIST OF ABBREVIATIONS	xiv
1 Introduction.....	1
1.1 Research Context	4
1.2 Research Aim and Objectives.....	8
1.3 Thesis Structure.....	9
1.4 Materials	14
1.4.1 Network Data	14
1.4.2 Soil Data.....	16
1.4.3 Weather Data	18
1.4.4 Data Processing Steps.....	19
1.5 References	20
2 Improving Pipe Failure Predictions: Factors Affecting Pipe Failure in Drinking Water Networks.....	25
2.1 Introduction	27
2.1.1 Drinking Water Pipe Failure	29
2.2 Pipe Intrinsic Factors Influencing Pipe Failure	31
2.2.1 Pipe Material	31
2.2.2 Joint Systems.....	34
2.2.3 Pipe Coating and Lining	37
2.2.4 Manufacturing Defects	38
2.2.5 Pipe Damage from Handling, Storage and Third Parties.....	40
2.2.6 Corrosion and Chemical Degradation	40
2.2.7 Pipe Age.....	43
2.2.8 Pipe Diameter	45
2.3 Environmental Factors Influencing Pipe Failure.....	46
2.3.1 Seasonality.....	47
2.3.2 Cold Temperature	48
2.3.3 Warm Temperature and Soil Movement.....	51
2.3.4 Other Soil Ground Movement Hazards	54
2.4 Operational Factors Influencing Pipe Failure	55
2.4.1 Internal Water Pressure	55
2.4.2 Previous Failures	56

2.5 Discussion	57
2.6 Conclusion	60
2.7 Author Contributions	61
2.8 References	61
3 Investigating the Effects of Operational Factors on Pipe Failure in Clean Water Networks.....	72
3.1 Introduction	74
3.2 Methodology	76
3.2.1 Data.....	76
3.3 Methods.....	78
3.3.1 Spatiotemporal Analysis.....	78
3.3.2 Joint Failures.....	79
3.3.3 Large Water Users, Flushing Events and Water source.....	79
3.4 Results and discussion	80
3.4.1 Spatiotemporal Relationship	80
3.4.2 Joint Failures.....	83
3.4.3 Proximity to Large Water Users	85
3.4.4 Flushing Events.....	86
3.4.5 Water Source	88
3.5 Conclusion	89
3.6 Author Contributions	90
3.7 References	90
4 Using Generalized Additive Models to Investigate the Environmental Effects on Pipe Failure in Clean Water Networks.....	96
4.1 Introduction	98
4.2 Methods.....	101
4.2.1 Data Assimilation	101
4.2.2 Statistical Analysis.....	104
4.2.3 Data Collection.....	106
4.3 Results.....	108
4.3.1 Iron.....	111
4.3.2 Asbestos Cement.....	113
4.3.3 Steel and Ductile Iron	115
4.3.4 Polyvinyl Chloride.....	117
4.3.5 Polyethylene.....	119
4.4 Discussion	120
4.4.1 Iron.....	120
4.4.2 Asbestos Cement.....	121
4.4.3 Steel and Ductile Iron	122
4.4.4 Polyvinyl Chloride.....	123
4.4.5 Polyethylene.....	124

4.5 Concluding Remarks.....	124
4.6 Author Contributions	126
4.7 References	126
5 An Evolution of Statistical Pipe Failure Models for Drinking Water Networks:	
A Targeted Review.....	134
5.1 Introduction	136
5.2 Heuristic Models	139
5.3 Physical Models.....	140
5.4 Statistical Models.....	140
5.4.1 Deterministic Models.....	141
5.4.2 Probabilistic Models	150
5.4.3 Machine Learning.....	167
5.5 Summary and Statistical Model Decision Aid.....	191
5.6 Conclusions	196
5.7 Author Contributions	198
5.8 References	198
6 Predicting the Risk of Pipe Failure Using Gradient Boosted Decision Trees	
and Weighted Risk Analysis	211
6.1 Introduction	213
6.2 Methodology	218
6.2.1 Decision Trees	218
6.2.2 Gradient Boosting	220
6.2.3 Weighted Risk Analysis.....	223
6.2.4 Data Curation	224
6.3 Results and Discussion.....	228
6.3.1 Receiver Operator Curve and Area Under the Curve.....	228
6.3.2 Confusion Matrix and Accuracy.....	231
6.3.3 Relative Variable Importance	234
6.3.4 Risk Mapping	236
6.3.5 Practical Considerations.....	237
6.3.6 Conclusion	240
6.4 Author Contributions	241
6.5 References	241
7 The Challenges of Predicting Pipe Failures in Clean Water Networks: A	
View from Current Practice.....	249
7.1 Introduction	251
7.2 Methodology	253
7.3 Results and Discussion.....	255
7.4 Challenges of Managing Data.....	256
7.4.1 An Incomplete Range of Correlated Variables	259
7.4.2 Data Silos and Data Sharing	260

7.5 Limitations of Pipe Failure Data	261
7.5.1 Coarse Recording the Time of Failure.....	261
7.5.2 A lack of High-Resolution Data	262
7.5.3 The Absence of Decommissioned Pipes.....	264
7.6 Complexities of Pre-Processing	266
7.6.1 Handling Poor Quality Data.....	266
7.6.2 Processing Incorrectly Placed Pipe Failures	268
7.6.3 Using Pipe Length and Previous Failures	271
7.7 Understanding Model Scalability.....	272
7.7.1 Scaling Geographic Extent.....	272
7.7.2 Scaling Dynamic Time-Dependent Variables	274
7.8 Future Opportunities and Challenges	275
7.9 Operational Practicality and Methodology of the Approach developed within this research	277
7.9.1 Operational Applicability of the Model	277
7.9.2 Practicality of the Spatiotemporal Scale	278
7.9.3 Variable Selection	279
7.9.4 Model accuracy	280
7.9.5 Comments.....	281
7.10 Key Findings	282
7.11 Study Limitations.....	283
7.12 Conclusion	284
7.13 Author Contributions	285
7.14 References	285
8 Summary of Findings and Implications of the Work	292
8.1 Introduction	292
8.2 Summary of Findings and Contribution.....	292
8.3 Implications of the Proposed Modelling Approach	300
8.3.1 Understanding Pipe Failures	300
8.3.2 Using Advanced Methods (Machine Learning).....	301
8.3.3 Determining Risk for Decision Makers	301
8.3.4 Reducing Leakage	302
9 Conclusion.....	303
9.1 Further Research Recommendations	306
9.2 References	308
Appendices.....	311
Appendix A Supplementary Material for Chapter 2.....	311
Appendix B - Supplementary Material for Chapter 3.....	312
Appendix C Supplementary Material for Chapter 4.....	313
Appendix D Supplementary Material for Chapter 6.....	321
Appendix E Supplementary Material for Chapter 7	332

LIST OF FIGURES

Figure 1-1 Top: Seasonal failure rates by material type. Bottom: Mean and mean minimum and maximum (data was taken from a weekly averaged data set) days air frost, temperature, rainfall, and soil moisture deficit in the region taken over 14 years between 2004 - 2018.....	6
Figure 1-2 Methodology flow diagram representing the sequence and interaction between the Objectives and Chapters.....	11
Figure 1-3 Simplified entity model showing the main water company datasets used in this research.	14
Figure 1-4 Simplified entity model showing the additional water company datasets used in this research.	16
Figure 1-5 Simplified entity model showing the soil dataset used in this research.	16
Figure 1-6 Simplified entity model for the Met Office weather data used in this research.....	19
Figure 2-1 Factors influencing the failure of drinking water pipes.....	30
Figure 2-2 Modes of failure (images courtesy of Anglian Water, 2018).	31
Figure 2-3 Approximate installation period of drinking water distribution pipe materials in the UK.	32
Figure 2-4 Typical joint systems and failures.....	36
Figure 2-5 Porosity and inclusion defects in a cast iron pipe.....	39
Figure 2-6 Cast iron pipe corrosion and chemical degradation.	42
Figure 2-7 Summary of introduction year and successive failure rates by age by material type (Utility Provider, 2018).....	45
Figure 2-8 Summary of failure rates by diameter size and pipe material type (Utility Provider, 2018).	46
Figure 2-9 Summary of seasonal failure rates by material type (Utility Provider, 2018).	48
Figure 2-10 Frost loading on pipe.....	51
Figure 2-11 Seasonal shrink swell effect in clay-rich soils (after Farewell et al., 2012, p. 10).	53
Figure 3-1 Spatial analysis of all pipe failures within 20 m of each other. Data from UK water company for the operational period from 2005 to 2018.	81

Figure 3-2 Temporal analysis of all pipe failures within 20 m of each other and up to 30 days separation time. Data from UK water company for the operational period from 2005 to 2018.....	83
Figure 3-3 Estimated pipe failure rates (bursts/1000 km/year) in DMAs with large users, separated by five common material type. The error bars show a 95 % confidence interval.....	86
Figure 3-4 Estimated pipe failure rates (bursts/1000 km/year) within two weeks of a flushing event and without flushing events, separated by five common material type. The error bars show a 95 % confidence interval.	88
Figure 3-5 Estimated pipe failure rates (bursts/1000 km/year) by water source, separated by five common material type. The error bars show a 95 % confidence interval.....	89
Figure 4-1 Generalized additive model - partial dependence plots showing the partial dependence of selected variables on Iron pipe failures.	112
Figure 4-2 Generalized additive model - partial dependence plots showing the partial dependence of selected variables on Asbestos Cement pipe failures.	114
Figure 4-3 Generalized additive model - partial dependence plots showing the partial dependence of selected variables on Steel and Ductile Iron pipe failures.....	116
Figure 4-4 Generalized additive model - partial dependence plots showing the partial effects of selected variables on Polyvinyl Chloride pipe failures. ...	118
Figure 4-5 Generalized additive model - partial dependence plots showing the partial dependence of selected variables on Polyethylene pipe failures. .	120
Figure 5-1 Taxonomy of Pipe Failure Models (adopted from Snider and McBean (2020))......	139
Figure 5-2 Illustrative example of a survival curve (adapted from Le Gat and Eisenbeis (2000)).	151
Figure 5-3 Pictorial representation of white-box, grey-box, and black-box models, showing how the greater physical knowledge exists, the better the interpretability of pipe failures (adapted from Rezania (2008)).	168
Figure 5-4 Illustrative example of <i>K</i> -means clustering in a two-dimensional space, depicting three clusters of the <i>n</i> data samples, based on Euclidean distance to the cluster centroid.	170
Figure 5-5 Simple Artificial Neural Network architecture illustrating only three network layers (adapted from (Wei et al., 2018))......	171

Figure 5-6 A simple one-dimensional illustration of an optimal linear hyperplane maximising the separation between two categories of data, with a single slack variable within the margin.....	177
Figure 5-7 An example of data partitioning by a classification decision tree. (i) shows the two-dimensional data space. (ii) shows the first condition for splitting the data by variable x_2 at 200, where disjoint region R_1 is ≤ 200 and disjoint region $R_2 > 200$. This creates two additional nodes for further conditions. (iii) shows the second condition for splitting data by variable x_1 at both 30 and 60 to create two more disjoint region R_3 and R_4 (adapted from Soguero-Ruiz et al. (2020)).	180
Figure 5-8 Depiction of the bagging process, showing the parallel trees returning classifiers averaged for the final model.....	181
Figure 5-9 Depiction of Boosting, like bagging, but using the residuals of the previous tree to improve the learning. The final model is the average of all trees.....	182
Figure 5-10 Prediction model decision aid to determine the most useful approach in different decision-making contexts, considering four important key factors, 1) the type of variables available, 2) the time interval for management interventions, 3) the spatial level of the model, and 4) the response type and level of inference required.	195
Figure 6-1 An example of data partitioning by a classification decision tree. (i) shows the two-dimensional data space. (ii) first condition for partitioning the data by variable x_2 at 200, where disjoint region R_1 is ≤ 200 and disjoint region $R_2 > 200$. (iii) second condition for partitioning data by variable x_1 at both 30 and 60 to create two more disjoint region R_3 and R_4 (Taken from Barton et al. (2022)).....	220
Figure 6-2 Gradient Boosting Tree ensemble model (Taken from Barton et al. (2022))	222
Figure 6-3 Test data accuracy, Receiver Operator Curve (ROC) curve with Area Under the Curve (AUC) measure of performance for all classification thresholds. The red line is the ROC curve, and the grey line represents the diagonal $y = x$ and a point where the curve is random.	229
Figure 6-4 Test data accuracy, calibration curve with Brier's score. The red line is the calibration curve; the grey line represents a perfect fit.	231
Figure 6-5 Relative Importance of the variables used in the global model.	235
Figure 6-6 Choropleth weighted risk map categorised using Natural Jenks. Risk is calculated as a measure of the probability of pipe failure and the consequence of damage to nearest property and water lost based on pipe diameter. The map represents a small section of the entire UK WDN.....	237

Figure 7-1 Systematic map of the themes and sub-themes explored in this study.	256
Figure 7-2 Pipe installations showing the removal of local soil conditions, and installation of pipe bedding and backfill (Photos taken by author).	264
Figure 7-3 Depiction of pipe failure data with missing historical data (left truncation), removed pipe failures (decommissioned pipes), and reported failures.	265
Figure 7-4 Illustrative example of a water distribution network with methods for the relocation of pipe failure data. Green – correctly geographically located pipe failures snapped to nearest pipe. Yellow – geographically misplaced pipe failure moved to the nearest pipe within 1 km. Red – failure shares no characteristics of any pipe within 1 km.	270
Figure B-1 Spatial analysis of all pipe failures within 1 km of each other. Data from UK water company for the operational period from 2005 to 2018. ..	312
Figure D-2 ROC curves and calibration curves for the original data, the over sampling, and the weighted classes.	326
Figure D-3 ROC curves and calibration curves for the original data, the over sampling, and the weighted classes.	328

LIST OF TABLES

Table 1-1 Summary of thesis structure and manuscript status.....	13
Table 2-1 Utility provider network data: pipe installation date, length, and number of failures by pipe material collected between 2005 and 2018.	29
Table 2-2 Summary of modern preferred drinking water mains materials (after Twort, Ratnayaka and Brandt, 2001).....	34
Table 2-3 Summary of the failure modes and mechanisms by material type. ..	57
Table 3-1 WDN data collected between 2005 and 2018	77
Table 3-2 Variables considered for the analysis, description, and type.....	78
Table 3-3 Percentage of failures within 1 m of a joint for five common pipe materials. Percentage calculated as a percentage of failures for that specific material.....	85
Table 4-1 Classification bands for pipe age.	101
Table 4-2 Classification bands for pipe diameter.....	102
Table 4-3 Variables considered for statistical analysis, description, and units.	102
Table 4-4 Summary of drinking water network used in the analysis.....	107
Table 4-5 Concurvity results for Generalized Additive Model Covariates.....	109
Table 5-1 Summary of the main deterministic pipe failure models.	148
Table 5-2 Summary of the main probabilistic models discussed.....	162
Table 5-3 Summary of the main machine learning pipe failure models.	187
Table 5-4 Advantages and disadvantages of deterministic, probabilistic and machine learning methods.....	191
Table 6-1 Consequences of failure, their weights and associated severity scores.	223
Table 6-2 WDN data collected between 2005 and 2018.	224
Table 6-3 Feature vector of variables selected for the pipe failure model, including description and data type.....	225
Table 6-4 Table of thresholds from training data, including True Positive, False Positive, False Negative, True Negative, Matthews Correlation Coefficient, percentage of pipe identified with a true failure, and the percentage of true failures predicted correctly. Only a few rows are presented for succinctness. Bold row text shows the final threshold chosen.	233

Table 7-1 Participants of the semi-structured interview.....	254
Table 7-2 Summary of key findings including challenges and priorities for future efforts to improve pipe failure models.282	
Table C-1 Shrink swell potential categories 1 – 6.	313
Table C-2 Hydrology of Soil Classes based on substrate hydrogeology (presence of aquifers, catchment and groundwater relationships, soil type and approximate depth to aquifer) (Boorman et al., 1995).	314
Table C-3 Generalized additive model summary for Iron pipes.....	314
Table C-4 Generalized additive model summary for Asbestos Cement pipes.	315
Table C-5 Generalized additive model summary for Steel & Ductile Iron pipes.	316
Table C-6 Generalized additive model summary for Polyvinyl Chloride pipes.	318
Table C-7 Generalized additive model summary for Polyethylene pipes.	319
Table D-1 Shrink swell potential categories 1 – 6.....	321
Table D-2 Hydrology of Soil Classes based on substrate hydrogeology (presence of aquifers, catchment and groundwater relationships, soil type and approximate depth to aquifer).....	322
Table D-3 Corrosion to iron categories and description.	323
Table D-4 Natural Soils Resource Institute (NSRI) clay vulnerability data set - Dominant clay vulnerability class between 1 – 9.	323
Table D-5 Accuracy Evaluation Metrics for Test Dataset	325
Table D-6 Accuracy Evaluation Metrics for Test Dataset	328
Table E-1 Interview Topic guide	333

LIST OF EQUATIONS

Eq.3-1 Dijkstra's Algorithm	78
Eq.4-1 Generalised Additive Model	104
Eq.4-2 Penalised Cubic Regression	105
Eq.5-1 Single-Variate Linear Model.....	141
Eq.5-2 Generalised Linear Model.....	144
Eq.5-3 Hazard Function	150
Eq.5-4 Cox-PH	151
Eq.5-5 Artificial Neural Networks.....	171
Eq.5-6 Evolutionary Polynomial Regression	174
Eq.5-7 Support Vector Machines.....	176
Eq.5-8 Decision Tree.....	179
Eq.5-9 Bagging.....	180
Eq.5-10 Boosting.....	181
Eq.6-1 Decision Tree.....	219
Eq.6-2 Gradient Boosting	221
Eq.6-3 True Positive Rate	228
Eq.6-4 False Positive Rate	228
Eq.6-5 Brier's Score	230
Eq.6-6 Classification.....	231
Eq.6-7 Matthews Correlation Coefficient	232
Eq.6-8 Relative Variable Importance	234

LIST OF ABBREVIATIONS

AC	Asbestos Cement
AMP	Asset Management Plan
ANN	Artificial Neural Networks
AUC	Area Under Curve
BWPHM	Bayesian Weibull Proportional Hazard Model
CoD	Coefficient of Determination
DAS	Distributed Acoustic Sensors
DBSCAN	Density-Based Clustering of Applications with Noise
DI	Ductile Iron
DIC	Deviance Information Criteria
DMA	District Metered Area
ELM	Extreme Learning Machine
EPR	Evolutionary Polynomial Regression
ESRI	Environmental Systems Research Institute
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GAM	Generalized Additive Model
GBT	Gradient Boosting Tree
GDPR	General Data Protection Regulation
GIS	Geographical Information System
GLM	Generalised Linear Model
GLMM	Generalised Linear Mixed Models
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GW	Groundwater
HOST	Hydrology of Soil Type
KNN	K-Nearest Neighbour
KM	Kilometres
LandIS	Land Information System
MCC	Matthews Correlation Coefficient

MCMC	Monte Carlo Markov Chain
MOGA	Multi-Objective Genetic Algorithm
MORECS	Meteorological Office Rainfall and Evaporation Calculation Systems
MW	Mixed Water
NHPP	Non-Homogenous Poisson Process
NMFD	National Mains Failure Database
NOPF	Number of Previous Failures
PDPs	Partial Dependence Plots
PE	Polyethylene
PH	Proportional Hazard
PVC	Polyvinyl Chloride
RD	Rainfall Deficit
REML	Restricted Maximum Likelihood
RF	Random Forest
RMSE	Root Mean Squared Error
ROC	Receiver Operating Curve
SDI	Steel and Ductile Iron
SMD	Soil Moisture Deficit
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
SW	Surface Water
TN	True Negative
TP	True Positive
TPR	True Positive Rate
UK	United Kingdom
UKWIR	UK Water Industry Research
UV	Ultraviolet
WDN	Water Distribution Network
ZINHPP	Zero Inflated Non-Homogenous Poisson Process
ZIP	Zero Inflation Poisson

1 Introduction

Water Distribution Networks (WDN) are critical infrastructures, providing clean water to millions of people. Water companies aim to provide a continuous and uninterrupted supply of clean drinkable water despite increasing pressures from climate change, an increasing population, changing water usage patterns and ageing infrastructure (Farrow et al., 2017). UK population growth is estimated to be greater than ten million people between 2006 and 2030 (Environment Agency, 2008), further increasing water demand by an estimated 35 % by 2050.

The Environment Agency has highlighted reducing water lost through pipe failure as a necessary means to address future water demand (Environment Agency, 2019). Ofwat, the Water Services Regulation Authority for England and Wales, supported moves to reduce loss through pipe failures by setting companies '*stretching performance commitments to reduce leakage*', and water companies responded in a letter from water UK to the Secretary of State in October 2018, committing to the delivery of a 50 % reduction in failures from 2017-18 levels (Ofwat, 2020a).

The UK has over 300,000 km of pipes across its regional WDNs, consisting of an antiquated asset base, including pipes that have been in continuous service since the Victorian era (Farrow et al., 2017). As a result, approximately 3 billion litres of clean water are leaked from the UK network daily due to pipe leaks and failures (Environment Agency, 2019). Pipe failures not only result in wasted water but also result in water discolouration, reduction in pipe pressure, damage to proximal infrastructure, sinkholes, sand wash out and cascading failures of other infrastructure provision, flooding of properties, disruption during repair and incurred repair costs (Barton et al., 2019).

Pipe failure reporting is both public and high profile. Ofwat monitors the performance of water companies through key performance indicators (KPIs) and customer experience surveys (water supply interruptions – hours per property) and reliability and availability (leakage – megalitres a day). In some instances,

water companies may fail to meet these Objectives. In such cases, the regulator may impose heavy fines (see Ofwat, 2020b), creating a demanding environment whereby failure detection and repair becomes a '24/7' necessity to achieve annual targets (Deadman, 2010).

Water companies attempt to mitigate the effects of pipe failures by managing the WDN by targeting damaged pipes for repair or replacement. The challenge is in managing assets in such a way that reduces pipe failures and optimises customer service and value through balancing the effects of performance, cost and risk (Deadman, 2010). Since its privatisation in 1989, the water utility sector has invested over £110 billion in pipe repair and replacement strategies (Institution of Civil Engineers, 2014). These investments are committed through Asset Management Plans (AMP), which aim to optimise investments through improved strategic and capital planning processes, carefully balancing water safety, quality, quantity failure rates, and costs for each maintenance option (Kakoudakis et al., 2018).

Plans are put in place targeting underperforming areas in the WDN, and proactive maintenance is introduced scheduled in advance after assessing a pipes tendency to fail. However, it has been estimated that typically only approximately 10 % of planned maintenance is completed, and 90 % of network maintenance remains reactive, due, in part, to the ageing infrastructure and the problems it causes. Reactive maintenance is less efficient than planned, proactive maintenance and is 50 % more expensive (Institution of Civil Engineers, 2014). There is, therefore, a desire in the water industry to become more proactive in managing assets. Reducing reactive maintenance to focus on planned maintenance will also help optimise investment planning and reduce inefficiencies and disruption to consumers (Povall and Bradley, 2014). Water companies are leveraging data-driven models to achieve proactive management, yet this approach hinges on the ability to predict pipe failures accurately. Understanding the conditions that lead to failure is therefore essential, and yet

this remains inherently very difficult since failures result from the complex interaction between many factors (Farrow et al., 2017).

Several approaches to predicting pipe failures have been developed, including heuristic, physical and statistical modelling techniques (Kleiner and Rajani, 2001; Nishiyama and Fillion, 2013; Scheidegger et al., 2015; St. Clair and Sinha, 2012; Wilson et al., 2017). Many water companies adopt a heuristic approach guided by expert judgement, detailed knowledge of the network and information on the number of previous failures per pipe (Snider and McBean, 2018). Heuristic models are often too simplistic to predict pipe failures accurately, the complexity of which requires more advanced methods (Snider and McBean, 2020; Tang et al., 2019). Physical models offer improvement and ordinarily describe the mechanisms contributing to pipe failure by analysing the pipe's stresses and strengths (Rajani and Kleiner, 2001). However, physical models are expensive since they require detailed data on pipe structural properties, internal and external loads, proximal factors, and material deterioration (Rajani and Kleiner, 2001), and gathering data can be time-consuming. The results of a physical model only represent specific pipes, which cannot be extrapolated to other areas of the network, limiting the use of the model. Statistical models are becoming more widely adopted since they can integrate a variety of data, are more cost-effective, can be extrapolated across entire networks and can also characterise uncertainty.

Many global studies implementing statistical models for assessing water leakage have been developed over the past 40 years, ranging from simple deterministic models to advanced machine learning (Robles-Velasco et al., 2020). Statistical models aim to predict pipe failures given historical data, inferring, from the data, a model that characterises the response based on several factors (Barton et al., 2020, 2019). Such models, and the accuracy of pipe failure predictions, are influenced by the unique prevailing risks of each WDN, including network characteristics, environmental factors, and operational management activities; therefore, no single model developed serves as a global solution. Some studies

have sought to be more diagnostic (Gould et al., 2011; Wols et al., 2019) and explain the reason for pipe failure but not adequately explain failure rates for decision-making. Other studies explicitly look at pipe failure predictions (Winkler et al., 2018) for decision-making, and some take this further to understand the risk of pipe failures (Francis et al., 2014). Finding an appropriate model is a complex task and assumes that the response is not random but is functionally related to factors in some form. Typically, models will include pipe age, material, length, and diameter as the basis for predicting failures, yet many other environmental and operational factors exist that explain the performance of the WDN. With the development of accurate datasets and data collection more frequently over the past few decades, pipe failure models have become more complex and often require more complex modelling techniques and more data (Robles-Velasco et al., 2021; Tang, 2018).

Whilst several models have been developed, numerous reported pipe failure models have achieved only poor predictive accuracy (Economou et al., 2012; Kabir et al., 2016; Mailhot et al., 2000; Tang et al., 2019; Wu and Liu, 2017), suggesting there remains an opportunity for improvement. Such improvements may proceed from exploring the different factors affecting pipe performance, using more advanced machine learning methods, developing alternative approaches to modelling scale, or deriving an improved understanding of how to determine best the risk of a pipe failure occurring.

1.1 Research Context

This research was undertaken through the DREAM CDT Industrial Doctorate Scheme and was funded by the Natural Environmental Research Council (NERC). A UK water company (requested anonymity) has acted as the industrial partner and provided the research case study and datasets. The UK water company manages a WDN that includes approximately 40,000 km of pipe in its WDN, distributed over a geographical area of 27,476 km². The WDN, like many in the UK, consists of disjointed pipes lengths of various materials. As the WDN has grown over time, many antiquated pipes remain in service today – some over

100 years old. The oldest pipes in service are iron (including cast and spun), followed by Asbestos Cement (AC), Steel and Ductile Iron (SDI), Polyvinyl Chloride (PVC) (collectively Unplasticised, Post Chlorinated and Molecular Orientated Polyvinyl Chloride) and Polyethylene (PE) (medium and high density). These materials comprise approximately 97 % of the network, whilst the remaining 3 % comprises rare materials (such as lead or glass-reinforced plastic) or are recorded as unclassified. Failure rates in these older pipe materials are typically higher annually than those constructed of newer materials, with iron pipes accounting for half of all failures and asbestos cement a quarter of all failures.

Typically, pipes fail when several conditions compound and stress the structural resiliency of an ageing pipe (Kleiner and Rajani, 2002). Environmentally, the UK WDN case study region is one of the driest in the UK, being prone to hot and cold temperature extremes, considerable fluctuations in soil moisture deficit and rapid changes in water temperature, which results in seasonal variations in pipe failure rates Figure 1-1. The area has varied geomorphology and a range of soil types that make the largely antiqued pipe network vulnerable to failure. Corrosive and compressible clay and peat soils corrode pipes, making them more susceptible to failure, and highly moveable clay soils suffer shrinkage in drought regions during the summer months. These combined conditions are more common in the region than in the rest of the UK. The most acidic soils are often highly sandy, where the freely draining sand means that leaked water drains freely through percolation, not being evident at the surface, which can conceal pipe failures for extended periods (Barton et al., 2020). Clay and peat soils can hold soil moisture during the winter, leading to impacts during freezing periods and increasing failure frequency. Operationally, management practices and water use can also lead to pipe failure. The network in question provides water to a large population of over four million people, and the constant stress on service requires continuous operational maintenance that can lead to pipe failures through excavation and construction.

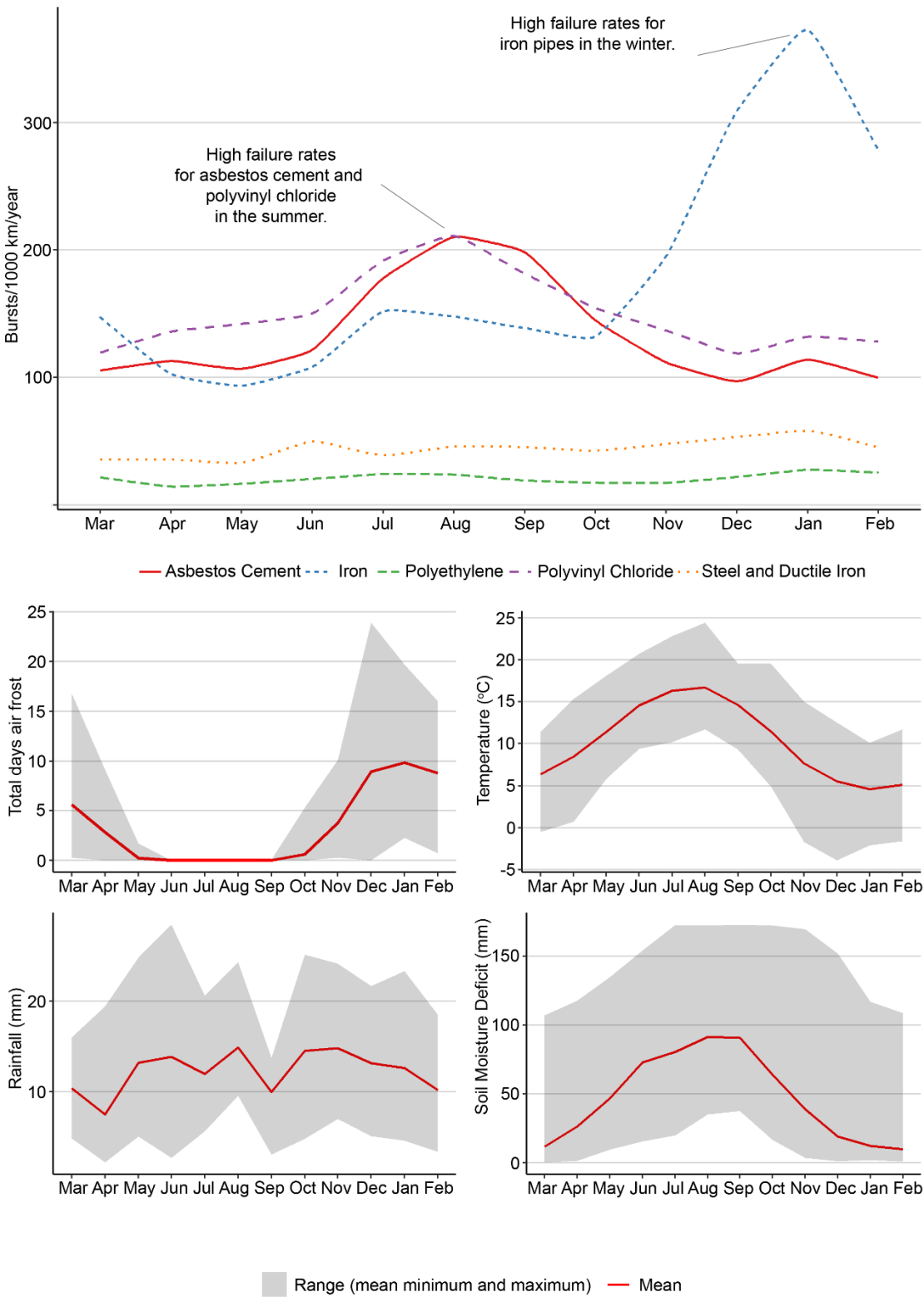


Figure 1-1 Top: Seasonal failure rates by material type. Bottom: Mean and mean minimum and maximum (data was taken from a weekly averaged data set) days air

frost, temperature, rainfall, and soil moisture deficit in the region taken over 14 years between 2004 - 2018.

Water companies undertake various levels of planning, defined here as short (monthly, inter-annual), medium (annual) or long-term (asset management period). Short-term planning is concerned with seasonal variations in pipe failures and applies to operational purposes. It is used to ensure enough resource is available for meeting the needs of reactively repairing the network. By contrast, medium or long-term asset management planning is used to decide on proactive rehabilitation, replacement, and maintenance of assets (Kleiner and Rajani, 2002). Decisions on asset management are complicated, and under normal operating conditions, several years can pass with the whole network in constant use, yet with only a small percentage of the network being replaced or rehabilitated due to budget constraints. Finding the highest risk pipes is therefore crucial, so water companies seek models to provide further important information (Deadman, 2010).

The UK water company used as a case study in this research currently uses statistical models to provide an accurate prediction of the number of annual asset failures, used to report leakage targets in the forthcoming reporting period. The model, which uses a Generalised Linear Model, provides more favourable estimates compared to the heuristic static approaches that failed to represent the variability in failures due to the influence of numerous operational, weather, and environmental factors. The motivation behind the original model development was to help meet leakage targets that were being missed, and consequently reduce the number of fines imposed on the water company. This instance was the first in the UK where a water company had aimed to use statistical models for reporting purposes.

However, reporting the number of annual asset failures is not useful for identifying pipes with a higher risk of failure. Therefore, the water company requires a more appropriate statistical model to predict pipe failures annually to inform asset management strategies that target areas of the network for pipe replacement,

rehabilitation, and maintenance. The motivation for this research is to reduce the predict pipe failures and the risk of pipe failure for small groups of pipes across the network, to help asset management approaches to derive long term key decisions, guided by evidence from past and predicted pipe failure information, minimising unnecessary expenditure and maximising returns on investment.

Ofwat incentivises and encourages the water industry to determine new, innovative ways of delivering customer services, including reducing leakage to improve serviceability (Ofwat, 2020c). Considering this, the UK water company wished to develop and implement advanced modelling techniques to help understand the potential for failures in the network. Alongside predicting pipe failures, water companies are also advised to adopt a risk-based approach to decision-making that guides replacements proportionate to the potential risk and promotes continual improvement. From this perspective, and in the interest of reducing pipe failures and enabling proactive maintenance, this research investigates pipe failures through data-driven techniques. It is anticipated that the modelling approaches proposed will enhance decision-making at a local level, facilitated through numerical and graphical outputs which report on the serviceability of the WDN and help meet Ofwat targets to reduce network failures and preserve water for future demand.

1.2 Research Aim and Objectives

The requirement for more accurate and representative methods for predicting pipe failure is essential to ensure reliable serviceability of WDNs and to reduce water loss to meet the needs of future water demand. Currently, there is scope in the UK to further develop the methodology of existing approaches.

For reliable pipe failure modelling this research aims to **investigate the causes of pipe failures in water distribution networks, develop a unique reliable methodology to predict the risk of pipe failure, and evaluate the pipe failure findings.**

The overall research aim is achieved through the following Objectives:

- **Objective 1:** Understand and characterise the various modes and mechanisms of pipe failure in WDNs.
- **Objective 2:** Investigate historical data from a UK water company to understand the main causes of failure and determine suitable variables for modelling.
- **Objective 3:** Compose and evaluate historical statistical pipe failure models to understand a potentially successful modelling approach.
- **Objective 4:** Develop a pipe failure modelling methodology that identifies pipe failures and the risk of failure and apply it to the UK WDN.
- **Objective 5:** Discuss and synthesise the challenges and limitations of modelling pipe failures with professionals currently in practice.

1.3 Thesis Structure

This thesis organises chapters presented as journal manuscripts. The manuscripts prepared here are each presented as an original piece of research addressing one or more of the Objectives outlined. Each contributes to original research and relates to the overarching aim. The research Objectives were conducted in sequence, such that the findings of one Objective helped inform subsequent Objectives. The manuscripts are included as presented when published, i.e., they may vary in presentation, heading style and terminology, particularly the use of either ‘drinking water networks’, ‘clean water networks’ and ‘water distribution networks’ used here interchangeably, but separately used in various chapter to suit the wishes of the Journal. Furthermore, the use of the word ‘generalized’, an American version of the spelling, rather than ‘generalised’ remains for this same reason. Acronyms may be provided more than once where some are repeated for each Chapter, and references may also be repeated as they are presented at the end of each chapter to keep with the thesis by paper structure. A short introduction is provided for each chapter, giving a brief overview, and setting each manuscript in the broader context of the thesis.

Appendices are provided where necessary at the end of each chapter. These include additional information and analysis completed to help shape the overall research.

Different methods have been used to address the research aim and Objectives. A research methodology flow diagram is presented in Figure 1-2. The diagram presents the sequence of Objectives and the relevant chapters.

Objective 1	Chapter 2: Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks	<ul style="list-style-type: none"> ▪ Reviewing existing literature on the factors that affect pipe failure. ▪ Understand the modes and mechanisms of pipe failure. ▪ Collect historical UK WDN variables relevant to pipe failures. ▪ Data cleaning, preprocessing and transformation.
Objective 2	Chapter 3: Investigating the effects of operational factors on pipe failures in clean water networks	<ul style="list-style-type: none"> ▪ Understand the UK WDN data by investigating how operational variables influence pipe failure. Use understanding of Objective 1 to transform data. ▪ Identify the most predictive operational variables to use when developing the pipe failure model.
	Chapter 4: Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks	<ul style="list-style-type: none"> ▪ Understand the UK WDN data by investigating how environmental variables influence pipe failure. Use understanding of Objective 1 to transform data. ▪ Identify the most predictive environmental variables to use when developing the pipe failure model.
Objective 3	Chapter 5: An evolution of statistical pipe failure models: a targeted review	<ul style="list-style-type: none"> ▪ Review existing literature on statistical pipe failure models. ▪ Identify the most appropriate model, response type, prediction interval and spatio-temporal scale given the available variables identified in Objective 2.
Objective 4	Chapter 6: Predicting the risk of pipe failure using gradient boosted decision trees and weighted risk analysis.	<ul style="list-style-type: none"> ▪ Using the modelling approach identified in Objective 3 and the variables from Objective 2, predict pipe failures using a machine learning approach. ▪ Develop suitable risk analysis to understand the risk of pipe failure. ▪ Use GIS to visualise areas of the network at high risk in the event of pipe failure.
Objective 5	Chapter 7: The challenges of predicting pipe failures in clean water networks: a view from current practice	<ul style="list-style-type: none"> ▪ Using the results and understanding from Objectives 3 and 4. Detail the challenges and limitations of pipe failure data and modelling. ▪ Understand the issues faced in industry by collecting qualitative data from semi-structured interviews with industry professionals.

Figure 1-2 Methodology flow diagram representing the sequence and interaction between the Objectives and Chapters.

Co-ordinating with the research Objectives outlined in Section 1.2, the first Objective 1 is presented in Chapter 2, which offers a literature review of the different factors affecting pipe failures. The review is supported by historical data from the UK WDN case study. This review contributes to a gap in knowledge established after collating the numerous factors and understanding the pipe failure complexities. Objective 2 deploys a quantitative method to analyse the UK WDN case study historical data. For brevity and to contribute to gaps in existing research. Chapter 3 evaluates the failure frequency of operational variables, some of which have not been used before, and applies a mixed-methods approach, including failure frequency, proximity analysis and Dijkstra's algorithm. Chapter 4 uses dynamic time-related environmental variables, a generalised additive model and partial dependence plots to understand failure frequency captured on a monthly interval. Objective 3, presented in Chapter 5 as a targeted review, evaluates the evolution and development of pipe failure models from simple single variate to complex data-driven machine learning methods. The findings are synthesised into a decision aid that proposes the most appropriate modelling strategies for different decision-making contexts. Historical developments and current thinking around various modelling approaches are discussed. Objective 4, presented in Chapter 6, details the quantitative data-driven machine learning methodology developed to predict the probability of pipe failure using a gradient boosting decision tree model. The probability of failure and weighted risk analysis is established from the consequence of the failure. Objective 5 investigates the challenges and limitations of pipe failure modelling in Chapter 7, from the perspective of industry professionals, achieved through a series of semi-structured interviews, and examines the views on the practicality of the model and the methodology developed. Chapter 8 provides a synthesis of findings from Chapters 2 - 7 and describes the wider context of the findings. Chapter 9 provides a conclusion and recommendations for further research. Chapters 8 and 9 are not prepared as a manuscript for publication.

This research comprises four published manuscripts and two draft manuscripts currently under review (Table 1-1). Author contributions to each manuscript are detailed within the publication. Neal Barton was the primary author for each manuscript undertaking conceptualisation, methodology, data curation, software, validation, formal analysis, investigation, visualisation and writing of all content, including drawing formal discussions and conclusions. Prof. Stephen Hallett and Dr Simon Jude contributed as academic supervisors, undertaking project administration, review, and editing. Dr Timothy Farewell provided additional advice and guidance for Chapters 2, 3 and 4 with expertise on soil-related geohazard modelling. Timothy Acland provided technical guidance for Chapter 2 with over 30 years of experience working in the UK water industry. Dr Trung Hieu Tran provided statistical modelling advice for Chapters 5 and 6 with expertise in machine learning and the water industry.

Table 1-1 Summary of thesis structure and manuscript status.

Cha.	Obj.	Title	Journal	Status
1	-	Introduction	Thesis only – no publication.	
2	1	Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. https://doi.org/10.1016/j.watres.2019.114926	Water Research	(2019) 164, 114926
3	2	Investigating the effects of operational factors on pipe failures in clean water networks.	Aqua	Currently under review
4	2	Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. https://doi.org/10.1038/s41545-020-0077-3	npj Clean Water	(2020) 3, 31
5	3	An evolution of statistical pipe failure modes for drinking water networks: a targeted review. https://doi.org/10.2166/ws.2022.019	Water Supply	(2022) 00, 0
6	4	Predicting the risk of pipe failure using gradient boosted decision trees and weighted risk analysis. https://doi.org/10.1038/s41545-022-00165-2	npj Clean Water	(2022) 5, 22
7	5	The challenges of predicting pipe failures in clean water networks: a view from current practice. https://doi.org/10.2166/ws.2021.255	Water Supply	(2021) 00, 0
8	-	Discussion – summary of findings and implications, and recommendations for further work. Conclusion	Thesis only – no publication.	

1.4 Materials

1.4.1 Network Data

The main infrastructure and pipe failure data use for all analysis was provided by the UK water company in two separate .shp files. The first data set contained the infrastructure information containing 40,000 km of pipe distributed over a geographical area of 27,476 km². The second data set contained the pipe failure information, with failure records collected between 2005 and 2018. The failure data contained numerous failures that were not geographically captured correctly, and often the pipe failure was displaced some distance from a nearby the pipe; a common occurrence in pipe failure datasets (Barton et al., 2021). Using ArcGIS, the water company relocated pipe failure records to the nearest matching pipe within the network, based on equivalent diameter size and material type to those recorded onsite. The snapping was first performed up to 3 m, and where pipe failures did not match a pipe section located within this 3 m distance, the process was repeated sequentially up to 1 km until a match was found. The two spatial datasets are shown in Figure 1-3.

Infrastructure (.shp)		Failures (.shp)	
Asset Number	int	Order Number	int
Material	text	X	dec
Length	int	Y	dec
Diameter band (See Table 1-3)	int	Failure Date	date
Age Band (See Table 1-4)	int	Material	text
Install Date	date	Diameter Band	int
DMA	text	Planning Type (Reactive or Proactive)	text

Figure 1-3 Simplified entity model showing the main water company datasets used in this research.

Table 1-2 Classification bands for pipe age.

Age Band	Range
1	< 1880
2	1880 to < 1900
3	1900 to < 1920
4	1920 to < 1940
5	1940 to < 1960
6	1960 to < 1980
7	1980 to < 2000
8	2000 to < 2021

Table 1-3 Classification bands for pipe diameter

Diameter band	Range
1	< 166 mm
2	166 to < 321 mm
3	321 to < 626 mm
4	≥ 626 mm

Additional datasets were also obtained from the water company to accompany the infrastructure and failure data when investigating the causes of failure. Four data sets were collected. The first contained information on water source zones including Groundwater (GW), Surface Water (SW) and Mixed Water (MW). The second dataset contained the location of large water consumers, based on a consumption greater than 10,000 m³ /year or 23 m³/day, as determined by the tariff, maximum daily demand charge. The third dataset contained information on the location of pipe joints in the network. The fourth dataset contained the location and times of recent flushing events. The four datasets are shown in Figure 1-4.

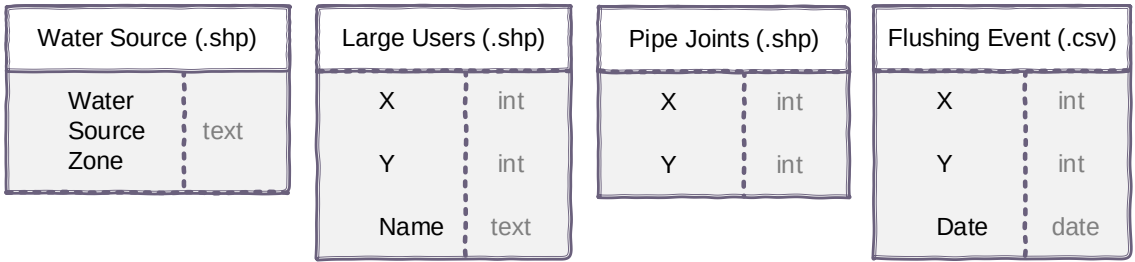


Figure 1-4 Simplified entity model showing the additional water company datasets used in this research.

1.4.2 Soil Data

Soil data was provided by Cranfield University’s National Soil Map and LandIS repositories. The data was presented in .shp files as 1:250,000 maps akin to a wide range of physical factors relating to soil, collected between 1939 and 1987 during an extensive field study campaign. The soil attributes are shown in

Figure 1-5.

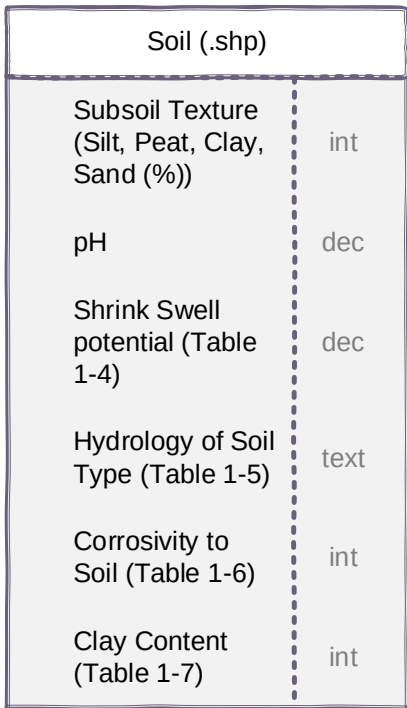


Figure 1-5 Simplified entity model showing the soil dataset used in this research.

Table 1-4 Shrink swell potential categories 1 – 6.

SSWELL Class	Shrink swell potential	Shrinkage (vol) %	Soil characteristics
Very low	1	< 3	Lithoskeletal, gravelly, sandy, light loamy or silty
Low	2	3 - 5	Medium loamy or medium silty unless derived from clay-with-flints/plateau drift
Moderate	3	5 - 12	Medium loamy or medium silty and derived from clay-with-flints or clayey on soft shales, Reddish Marl, clay and sand, non-swelling clay, or any till or head deposits
High	4	12 -15	Clayey and on swelling clay, clay-with flints/plateau drift or glaciolacustrine clay
Very High	5	> 15	Clayey and on brownish swelling clay
High	6	12 - 15*	Alluvial clay or peat – very high SSWELL. *Potential that is not realised unless effective drainage is installed to at least 2m depth.

Table 1-5 Hydrology of Soil Classes based on substrate hydrogeology (presence of aquifers, catchment and groundwater relationships, soil type and approximate depth to aquifer) (Boorman et al., 1995).

1	Weakly consolidated, microporous, by-pass flow uncommon.
2	Weakly consolidated, macroporous, by-pass flow uncommon.
3	Strongly consolidated, non or slightly porous, by-pass flow common.
4	Unconsolidated, macroporous by-pass flow very uncommon.
5	Unconsolidated, macroporous by-pass flow common.
6	Slowly Permeable.
7	Impermeable (hard).
8	Impermeable (soft)
9	Eroded Peat.
10	Raw Peat.

Table 1-6 Corrosion to iron categories and description.

Corrosivity Class	Description
1	Non-aggressive
2	Slightly aggressive
3	Moderately aggressive
4	Highly aggressive
5	Very highly aggressive

Table 1-7 Clay vulnerability data set - Dominant clay vulnerability class between 1 – 9. Indicates the level of movement and potential ground subsidence.

Clay Class	Description
1	Extremely low
2	Very low
3	High
4	Low
5	Medium low
6	Medium
7	Medium High
8	Very High
9	Extremely High

1.4.3 Weather Data

Weather and climate data was provided by the Met Office Rainfall Calculation System (MORECS). The data provides weekly estimates of temperature, Soil Moisture Deficit (SMD) and rainfall at a spatial resolution of 40 km. The data was provided in a .csv file format, with a .shp of the grid squares and unique IDs that could be related to the data. Further data was collected via a web scrape from the Met Office National Climate Information centre on the monthly number of days

with sir frost (minimum temperature below zero) on a regional scale. The weather data was collected for the same period as the failure records, between 2005 and 2018. The final datasets are shown in Figure 1-6.

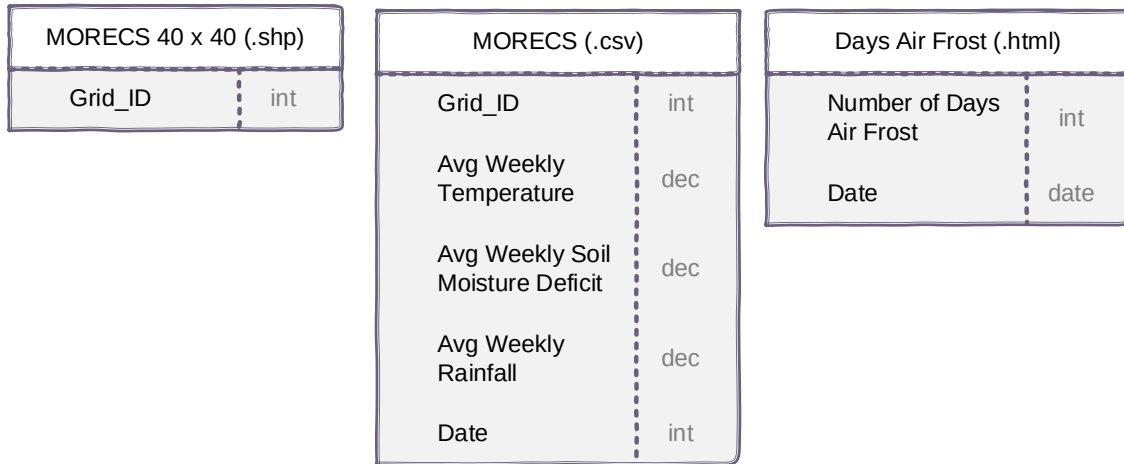


Figure 1-6 Simplified entity model for the Met Office weather data used in this research.

1.4.4 Data Processing Steps

The following details the data processing steps taken to set up the basic dataset to work from.

- 1) The following steps were performed in ArcGIS:
 - a) The infrastructure data was intersected with the soil data to create the pipe lengths with various infrastructure and soil characteristics data.
 - b) The failure data, water source data and MORECS Grid_ID were attributed to each pipe length.
 - c) A 1km grid that sat explicitly inside the MORECS 40x40 km grid squares with a Unique ID were attributed to each pipe length.
 - d) Pipes with similar characteristics were summarised, constrained by the 1 km grid to create groups of pipes.
- 2) R Data processing
 - a) The network was replicated over the full weekly data range to include the weather data.

- b) The MORECS weather data was added to each pipe length for the correct week date.
- c) Data transformations were created, including the following:
 - i) For Chapter 4 - Temperature change over one week, temperature change over two weeks, SMD change over one week, SMD change over two weeks; and
 - ii) For Chapter 6 – Weeks Temperature > 20°C, Weeks Temp < 5°C, Weeks SMD > 120 mm and number of previous failures.

Once the data set was created, the time scale could be adjusted, summarising the weekly data accordingly. The individual operational data sets for flushing events, large users and pipe joints were processed separately for the single variate analysis presented in Chapter 3. Although the same data was used, the methodology was different. Please see Section 3.3 for specific details on these.

1.5 References

- Barton, N.A., Farewell, T.S., Hallett, S.H., 2020. Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *npj Clean Water* 3, 31. <https://doi.org/10.1038/s41545-020-0077-3>
- Barton, N.A., Farewell, T.S., Hallett, S.H., Acland, T.F., 2019. Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Res.* 164, 114926. <https://doi.org/10.1016/j.watres.2019.114926>
- Deadman, C., 2010. *Strategic Asset Management: The quest for utility excellence*. Matador, Leicester, UK.
- Economou, T., Kapelan, Z., Bailey, T.C., 2012. On the prediction of underground water pipe failures: zero inflation and pipe-specific effects. *J. Hydroinformatics* 14, 872–883. <https://doi.org/10.2166/hydro.2012.144>

- Environment Agency, 2019. Escaping the jaws of death: ensuring enough water in 2050 [WWW Document]. URL <https://www.gov.uk/government/speeches/escaping-the-jaws-of-death-ensuring-enough-water-in-2050>
- Environment Agency, 2008. Water resources in England and Wales - current state and future pressures.
- Farrow, J., Jesson, D., Mulheron, M., Nensi, T., Smith, P., 2017. Achieving zero leakage by 2050: the basic mechanisms of bursts and leakage, UK Water Industry Research Limited. London.
- Francis, R.A., Guikema, S.D. and Henneman, L. (2014) 'Bayesian Belief Networks for predicting drinking water distribution system pipe breaks', Reliability Engineering & System Safety, 130 Elsevier, pp. 1–11.
- Gould, S.J.F., Boulaire, F.A., Burn, S., Zhao, X.L., Kodikara, J.K., 2011. Seasonal factors influencing the failure of buried water reticulation pipes. Water Sci. Technol. 63, 2692–2699. <https://doi.org/10.2166/wst.2011.507>
- Institution of Civil Engineers, 2014. The State of The Nation Infrastructure 2014. Inst. Civ. Eng.
- Kabir, G., Tesfamariam, S., Loepky, J., Sadiq, R., 2016. Predicting water main failures: A Bayesian model updating approach. Knowledge-Based Syst. 110, 144–156. <https://doi.org/10.1016/j.knosys.2016.07.024>
- Kakoudakis, K., Farmani, R., Butler, D., 2018. Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. J. Hydroinformatics 20, 1191–1200. <https://doi.org/10.2166/hydro.2018.152>
- Kleiner, Y., Rajani, B., 2001. Comprehensive review of structural deterioration of water mains: statistical models. Urban Water 3, 131–150. [https://doi.org/10.1016/S1462-0758\(01\)00033-4](https://doi.org/10.1016/S1462-0758(01)00033-4)

- Kleiner, Y., Rajani, B., 2002. Forecasting Variations and Trends in Water-Main Breaks. *J. Infrastruct. Syst.* 8, 122–131. [https://doi.org/10.1061/\(ASCE\)1076-0342\(2002\)8:4\(122\)](https://doi.org/10.1061/(ASCE)1076-0342(2002)8:4(122))
- Mailhot, A., Pelletier, G., Noël, J.-F., Villeneuve, J.-P., 2000. Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application. *Water Resour. Res.* 36, 3053–3062. <https://doi.org/10.1029/2000WR900185>
- Nishiyama, M., Fillion, Y., 2013. Review of statistical water main break prediction models. *Can. J. Civ. Eng.* 40, 972–979. <https://doi.org/10.1139/cjce-2012-0424>
- Ofwat, 2020a. Leakage [WWW Document]. URL <https://www.Ofwat.gov.uk/households/supply-and-standards/leakage/> (accessed 9.25.20).
- Ofwat, 2020b. PN 22/18: Thames Water's failure to tackle leakage results in £65m package for customers [WWW Document]. URL <https://www.Ofwat.gov.uk/pn-22-18-thames-waters-failure-tackle-leakage-results-65m-package-customers/> (accessed 1.5.21).
- Ofwat, 2020c. Innovation in the water sector [WWW Document]. URL <https://www.Ofwat.gov.uk/regulated-companies/innovation-in-the-water-sector/>
- Povall, T., Bradley, G., 2014. Getting Ready for AMP6: The Opex Challenge 8.
- Rajani, B., Kleiner, Y., 2001. Comprehensive review of structural deterioration of water mains: Physically based models. *Urban Water* 3, 151–164. [https://doi.org/10.1016/S1462-0758\(01\)00032-2](https://doi.org/10.1016/S1462-0758(01)00032-2)
- 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. *Reliab. Eng. Syst. Saf.* 196, 106754. <https://doi.org/10.1016/j.ress.2019.106754>

- Robles Velasco, A., Muñuzuri, J., Onieva, L., Rodríguez Palero, M., 2021. Trends and applications of machine learning in water supply networks management. *J. Ind. Eng. Manag.* 14, 45. <https://doi.org/10.3926/jiem.3280>
- Scheidegger, A., Leitão, J.P., Scholten, L., 2015. Statistical failure models for water distribution pipes – A review from a unified perspective. *Water Res.* 83, 237–247. <https://doi.org/10.1016/j.watres.2015.06.027>
- Snider, B., McBean, E.A., 2020. Watermain breaks and data: the intricate relationship between data availability and accuracy of predictions. *Urban Water J.* 17, 163–176. <https://doi.org/10.1080/1573062X.2020.1748664>
- Snider, B., McBean, E.A., 2018. Improving Time-to-Failure Predictions for Water Distribution Systems using Gradient Boosting Algorithm, in: 1st International WDSA / CCWI 2018 Joint Conference. Kingston, Ontario.
- St. Clair, A.M., Sinha, S., 2012. State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models! *Urban Water J.* 9, 85–112. <https://doi.org/10.1080/1573062X.2011.644566>
- Tang, K., 2018. Statistical Modelling Approaches to Improve Risk Assessments for United Kingdom's Water Infrastructure. Cranf. Univ.
- Tang, K., Parsons, D.J., Jude, S., 2019. Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliab. Eng. Syst. Saf.* 186, 24–36. <https://doi.org/10.1016/j.res.2019.02.001>
- Wilson, D., Fillion, Y., Moore, I., 2017. State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water J.* 14, 173–184. <https://doi.org/10.1080/1573062X.2015.1080848>
- Wols, B.A., Vogelaar, A., Moerman, A., Raterman, B., 2019. Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. *Water Supply* 19, 404–416. <https://doi.org/10.2166/ws.2018.085>

Wu, Y., Liu, S., 2017. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water J.* 14, 972–983.
<https://doi.org/10.1080/1573062X.2017.1279191>

2 Improving Pipe Failure Predictions: Factors Affecting Pipe Failure in Drinking Water Networks

Objective 1	Chapter 2: Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks	<ul style="list-style-type: none">▪ Reviewing existing literature on the factors that affect pipe failure.▪ Understand the modes and mechanisms of pipe failure.▪ Collect historical UK WDN variables relevant to pipe failures.▪ Data cleaning, preprocessing and transformation.

The following chapter reviews the modes and mechanisms of pipe failure in clean water networks and fulfils Objective 1. This chapter reviews the existing literature to understand the major factors affecting pipe failure frequency, whilst data collected and curated from the UK WDN case study supports the findings. The following chapter has been taken directly, without further edits, from the following review article published in water research:

Barton, N.A., Farewell, T.S., Hallett, S.H., Acland, T.F., 2019. Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. Water Res. 164, 114926.
<https://doi.org/10.1016/j.watres.2019.114926>

Improving Pipe Failure Predictions: Factors Affecting Pipe Failure in Drinking Water Networks

Neal Andrew Barton^a, Timothy Stephen Farewell^{a*}, Stephen Henry Hallett^a, and Timothy Francis Acland^b

^a School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

^b 5D, Kettering, Northamptonshire, NN15 5AW, United Kingdom.

**Corresponding author: Email: t.s.farewell@cranfield.ac.uk; Tel: +441234752978*

Abstract

To reduce leakage and improve service levels, water companies are increasingly using statistical models of pipe failure using infrastructure, weather, and environmental data. However, these models are often built by environmental data scientists with limited in-field experience of either fixing pipes or recording data about network failures. As infrastructure data can be inconsistent, incomplete, and incorrect, this disconnect between model builders and field operatives can lead to logical errors in how datasets are interpreted and used to create predictive models. An improved understanding of pipe failure can facilitate improved selection of model inputs and the modelling approach. To enable data scientists to build more accurate predictive models of pipe failure, this paper summarises typical factors influencing failure for 5 common groups of materials for water pipes: 1) cast and spun iron, 2) ductile iron, 3) steel, 4) asbestos cement, 5)

polyvinyl chloride (PVC) and 6) polyethylene (PE) pipes. With an improved understanding of why and how pipes fail, data scientists can avoid misunderstanding and misusing infrastructure and environmental data and build more accurate models of infrastructure failure.

Key Words: Pipe failure; Water supply; infrastructure planning; Environment; Soil.

Highlights:

- Pipe failure models can be improved with an understanding of failure mechanisms.
- Examples of failure data from the literature is supported by a large water utility's failure data.
- Pipe failure modes are summarised for common pipe materials.
- Environmental, operational & pipe intrinsic factors impacting failure are discussed.
- Summaries are provided for key failure mechanisms for common pipe materials.

2.1 Introduction

Across the UK water network, approximately 22 % of all treated drinking water is lost through water pipe failure (Farrow et al., 2017). The UK water industry regulator challenges water companies to reduce leakage and service interruptions, and to increase sustainability as future demands continue to grow. To reduce UK failure rates (currently approximately 170 bursts/1000 km/year), water utilities manage assets proactively. An area of interest for asset management is the development of quantitative tools, such as physical and statistical models to prioritise pipe repair and replacement. Statistical models correlate historic failures with observed conditions to predict future failures (Clark et al., 1982; Pelletier et al., 2003; Rajani and Kleiner, 2001).

Understanding modes and mechanisms of pipe failure from historical data is useful for predictive modelling as it can prevent illogical errors which arise from purely data driven approaches, which in turn can result in unrealistic assumptions or additional work through measuring the wrong thing. Complications can arise where data limitations are present (for example where a lack of awareness prevails as to the importance of data collection), or from common data handling practices which can omit records and introduce bias, or sparse data (Lin et al., 2015). As a result, data errors become more common (Rajani et al., 2012; Asnaashari et al., 2013; Scheidegger et al., 2013). Existing research provides a foundation for developing an understanding of the complex interactions between factors influencing pipe failure, which form mechanisms for failure (Pelletier et al., 2003; Rajeev et al., 2014; Rezaei et al., 2015). Together, this understanding provides a strong logical foundation for data and infrastructure scientists to develop such statistical models.

With a focus on assisting environmental and infrastructure data scientists to improve predictive modelling, this paper summarises extensive existing research on factors influencing multiple mechanisms of water pipe failure from the international literature. The findings are supported by contributions from industry professionals and network data supplied by a UK water distribution network operator (referred to as the utility provider). Information has been provided in Table 2-1, offering a foundation to apply future evaluation of individual pipe material failure for statistical modelling. We discuss each of the main pipe materials, constituting the majority of water network in the UK, namely: Iron (including cast and spun), Ductile Iron (DI), Steel, Asbestos Cement (AC), Polyvinyl Chloride (PVC) (collectively Unplasticised, Post Chlorinated and Molecular Orientated Polyvinyl Chloride) and Polyethylene (PE) (medium and high density).

Table 2-1 Utility provider network data: pipe installation date, length, and number of failures by pipe material collected between 2005 and 2018.

Material	Installation range	Total Network length (km)	Total No. of Pipe Failures
I	1881 to 1921	11,735	26,600
AC	1920 to 1941	7,259	14,053
PVC	1960 to 2001	6,126	11,942
PE	1981 to present	10,538	4,356
SDI	1960 to present	1,902	1,067
Total		37,560	58,018

2.1.1 Drinking Water Pipe Failure

Industry professionals and published literature use the terms leak, burst or failure when a pipe breaks and water is released. These terms are often synonymously used, and this paper has adopted the term failure throughout. Pipe failures in this context represents all pipe breaks and leaks that occur and require repair (Laucelli et al., 2014).

There are many different modes and mechanisms for pipe failure. Complex relationships between factors and their relative contributions to the failure mechanism are unique for each pipe material and geographic region (Gould et al., 2013; Rajeev et al., 2014). Factors can be categorised to three groups: 1) pipe-intrinsic, 2) operational and 3) environmental. Figure 2-1 presents the major factors affecting water pipe failure.

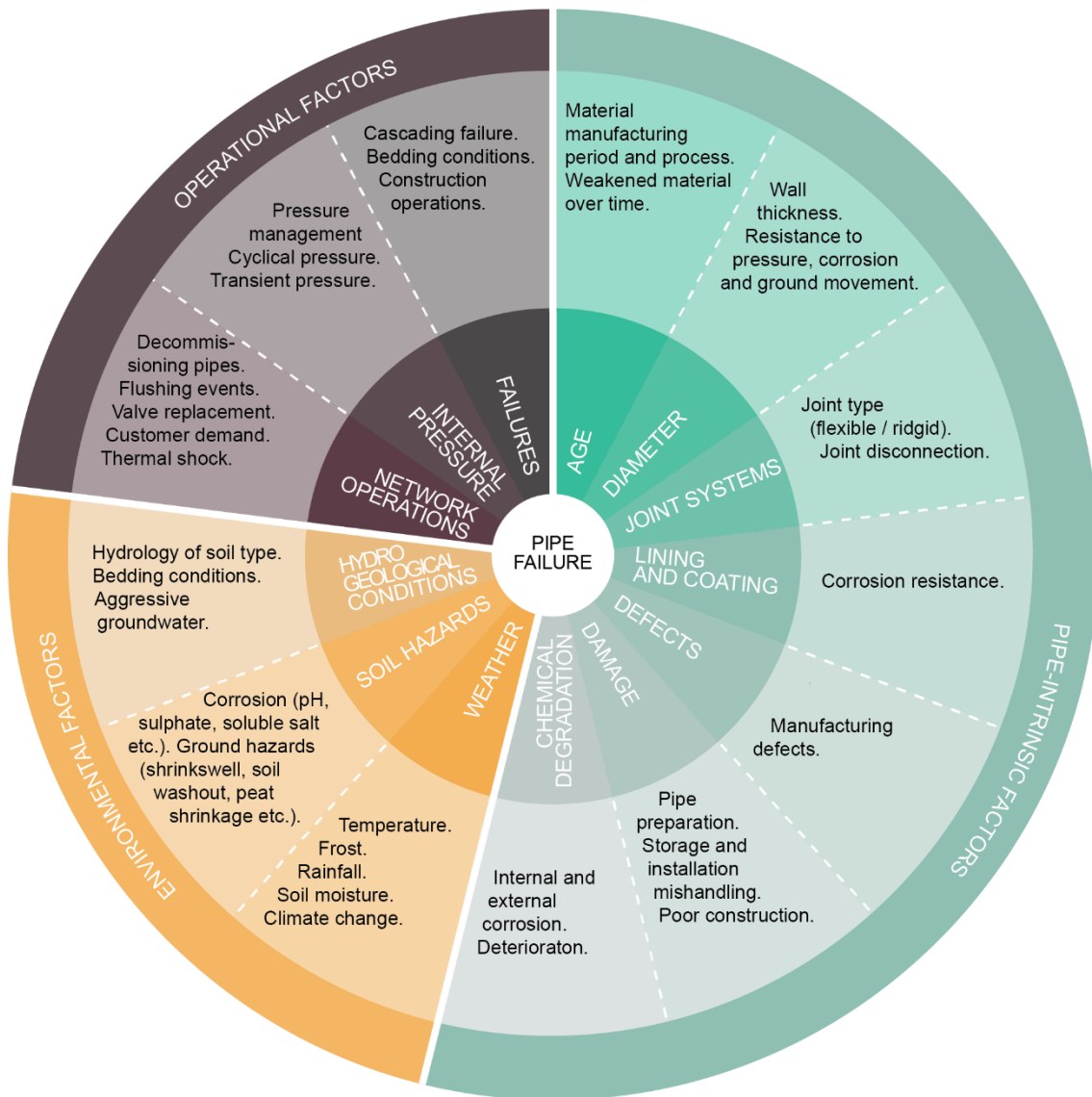


Figure 2-1 Factors influencing the failure of drinking water pipes.

Typical water pipe failure modes include circumferential break, longitudinal split, joint failure, and holes (both blowouts and pinhole leaks) (Farrow et al., 2017). Failure modes are associated with differing forces acting on the pipe. A typical circumferential break is often caused by tensile forces (soil movement or thermal expansion and contraction) and loading (heavy traffic) forces. Longitudinal splits are often caused by transverse and radial forces (e.g., internal water pressure) possibly in conjunction with a pre-existing defect acting as a point of weakness.

Joint failures are typically caused by tensile or compressive forces, while holes are typically caused by radial forces in conjunction with corrosion (Hu and Hubble, 2007; Makar et al., 2001). Figure 2-2 shows different failure modes acting *in situ*.

The main factors influencing pipe failure; 1) pipe intrinsic, 2) environmental and 3) operational are discussed in detail in the following section. Readers seeking a general summary of the important factors influencing failure in each pipe material are directed to Table 2-3 in the Discussion.



Circumferential break
on an asbestos cement pipe.



Longitudinal split
on a polyvinyl chloride pipe.



Corrosion pin hole
on an iron pipe.



Joint failure (disconnection or gasket
failure) on an asbestos cement pipe.

Figure 2-2 Modes of failure (images courtesy of Anglian Water, 2018).

2.2 Pipe Intrinsic Factors Influencing Pipe Failure

2.2.1 Pipe Material

Cast iron is one of the oldest pipe materials in the UK, with some pipes dating back even to the mid-1800s (Figure 2-3). Iron pipes are rigid and had three main

manufacturing methods, in sequence: horizontal pit casting, vertical pit casting (cast iron) and centrifugal spinning (spun iron), with each method improving on the former. Cast iron and spun iron pipes are considered together in this review as they have similar failure mechanisms but are differentiated from the more resilient ductile iron. Ductile iron (DI) and steel were introduced in the late 1950s as alternatives to iron pipes. DI is also manufactured using the centrifugal spun method, but to a stricter manufacturing standard. The material introduces magnesium into the alloy to change the graphite within the microstructure to spheres instead of flakes, producing a tougher material as compared to cast and spun iron. Steel is stronger and more ductile than cast iron, but has a lower tensile strength, and resistance to corrosion (requires more maintenance *in situ* such as cathodic protection) than DI, meaning that DI is preferred for use in diameter pipes typically between 300 – 800 mm in diameter. However, steel is cheaper than DI and its welded joints offer many advantages in high pressure pipes and where working area is limited. As a result, steel is the preferred material for large pipes > 800 mm in diameter (Ductile Iron Pipe Research Association, 1984; Ruchti, 2017; Twort et al., 2001).

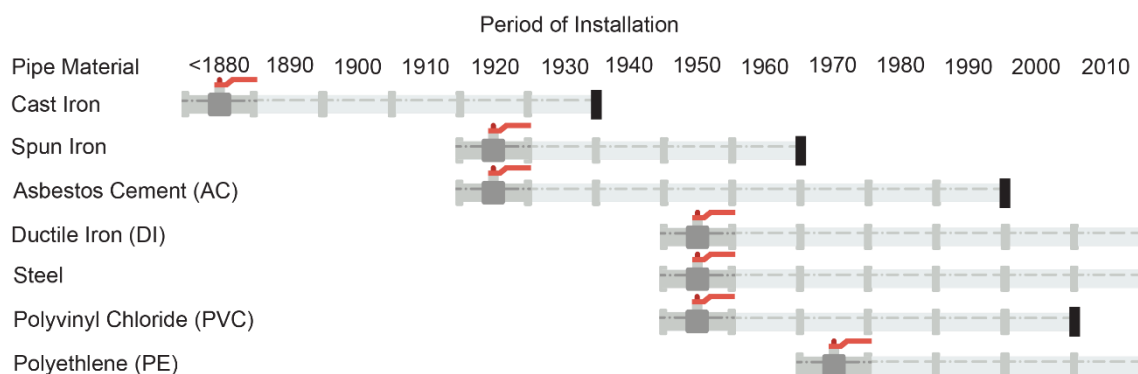


Figure 2-3 Approximate installation period of drinking water distribution pipe materials in the UK.

UK water companies introduced asbestos cement (AC) pipes as early as 1929 (Van Erp et al., 2015). However, the most prolific period of AC installation was between 1950s and 1960s. AC was cheap to manufacture, cheap to operate (low

frictional resistance) and generally resistant to corrosion. However, corrosion did occur in sulphate rich soils or where acidic ground water was present. Al-Adeeb and Matti, (1984) also reported leaching of free lime in pipes that conveyed very soft water. However, AC is a rigid and brittle material and less flexible than DI and steel (Mordak and Wheeler, 1988) so it is less resilient to ground movement. In 1986, manufacturing of AC pipe ceased in the UK due to the negative health perceptions surrounding asbestos in building materials (Twort et al., 2001). By the late 1980s, AC pipes accounted for approximately 11 % of the UK's water distribution network (Mordak and Wheeler, 1988).

PVC was introduced in the late 1950s and provided a corrosion resistant and flexible alternative to AC. PVC became popularized in the 1970s, however, during this period the PVC manufacturing process produced low degrees of gelation (where plasticizers diffuse into PVC particles) which resulted in low grade PVC pipes with a low toughness. PVC has rapidly increased in use over the last decade due to improved manufacturing processes (in particular the gelation process), low manufacturing costs, corrosion resistance, and ease of assemblage (Beuken et al., 2012; Bruaset and Sægrov, 2018; EL-Bagory and Younan, 2016; Guoquan and Yiaoting, 1991). Today, PVC accounts for some 13 % of the UK network. By the 1980s, PE was also widely used alongside PVC, but has eventually replaced PVC as it withstands higher pressure and lasts longer. PE remains the material of choice today (Farrow et al., 2017; Ruchti, 2017).

The choice of materials used throughout the timeline is subject to technical considerations, such as material availability, cost, installer experience and skills, ground condition and technician preference to name a few. The extent of pipe material installation in recent years is presented in Table 2-2.

Table 2-2 Summary of modern preferred drinking water mains materials (after Twort, Ratnayaka and Brandt, 2001)

Pipe diameter size (mm)	Material used	Explanation
≤ 50	PE	Cost effective at small diameters, can be joined easily above ground, and placed into narrow trenches.
> 50 - < 300	PE or PVC DI*	PE and PVC are more cost effective at small diameters, can be joined easily above ground and placed into narrow trenches. Where pressure and diameter increase DI may be favourable due to cost.
> 300 - < 800	PVC or DI Steel*	PVC or DI are predominantly used and more cost effective in middle diameter pipes. PVC is cheap, flexible, and resistant to corrosion. DI is strong and ductile and suitable for many ground conditions. The design and construction is simple due to compatible range of easy assemble fittings.
≥ 800	Steel DI*	Steel is mainly used for trunk mains or high-pressure mains, since welded joints are stronger, provide longitudinal strength and can easily fit in narrow corridors. DI is only used if the price is competitive or no skilled welders are available.

* Less commonly used

2.2.2 Joint Systems

Joint systems are integral to pipe networks, and are considered synonymously with the pipe (Trew et al., 1995). Joint systems can be similar for different pipe materials (such as the bolted mechanical) or can be specific to material type (electrofusion, used only in PE pipes). Today, the most common joints include the spigot and socket, bolted mechanical, flanged, butt welded and push fit (Figure 2-4). Joint systems are typically sealed with a gasket to ensure a tight seal. Joint systems can be integral (built as part of the pipe) or non-integral (connected separately to two pipe ends). They can be rigid (using flanged or

mechanical bolted joints which offer little flexibility) or flexible (push fit or spigot and socket joints which allow the pipe a small margin of movement).

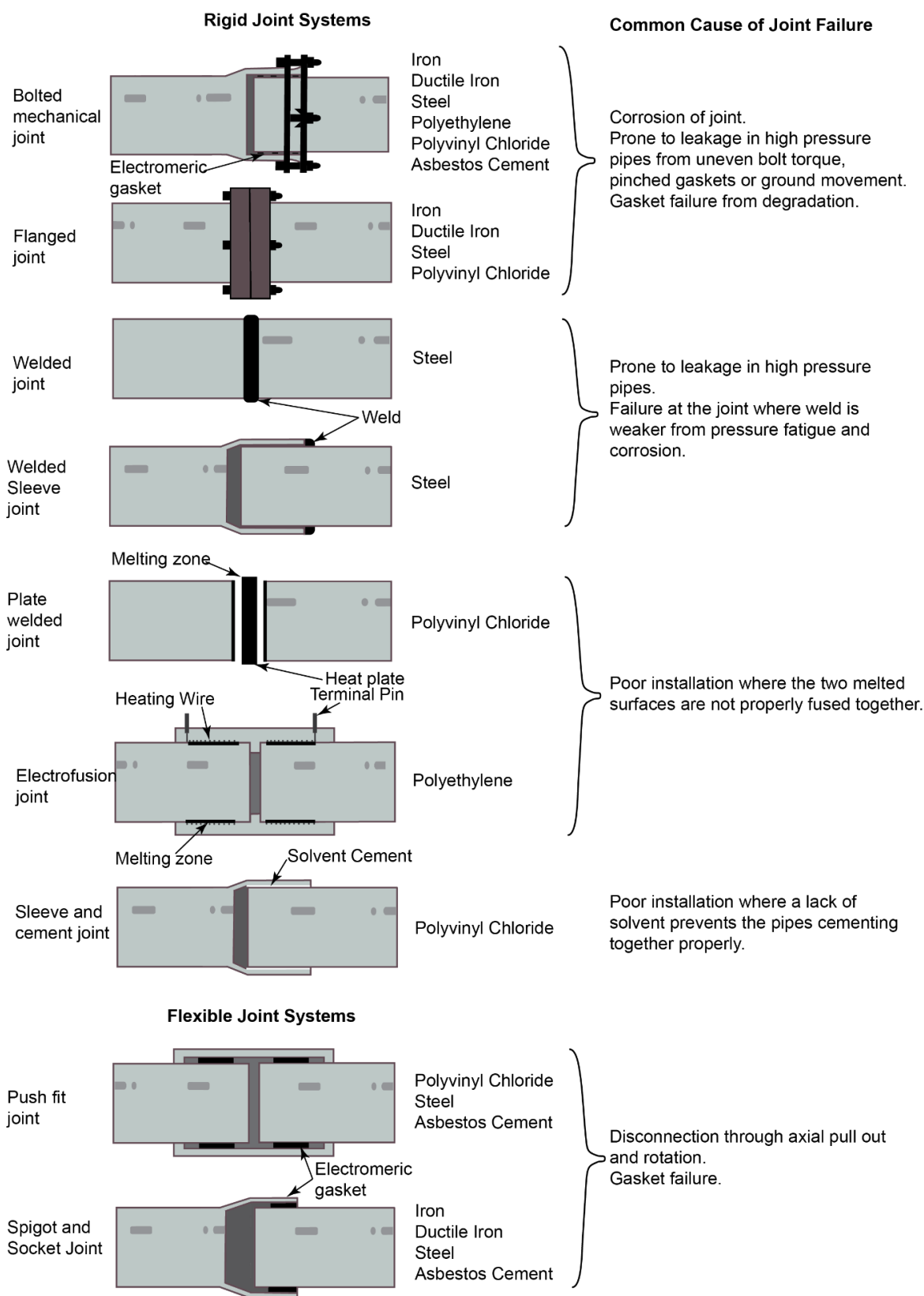


Figure 2-4 Typical joint systems and failures.

Joint failures include 1) joint leaking, 2) joint fracture, 3) disconnection or 4) gasket failure. Joint leaking can be a result of either poor installation, joint type, or ground movement after installation. Typically rigid joints leak as a result of poor installation such as uneven bolt torque, or pinched gaskets, thermal expansion and corrosion of bolts or connection surfaces (Ruchti, 2017; Twort et al., 2001). However, rigid joints are more susceptible to leakage and joint fracture as a result of ground movements, where multi-directional tensile forces exert stress on the pipe (Farrow et al. 2017). Flexible joints are designed to withstand small movement, and are favourable in areas of high ground movement, but multi-directional ground movement or poor installation can result in flexible joint disconnection, (i.e. forcing a straight joint onto two angled pipes). Dingus et al., (2002) and Burn et al., (2005) reported that 15 % and 16 % (average) respectively of all PVC pipe failures were due to joint failure. Kirby (1981) reported that over insertion and angularity of the pipe accounts for 12.4 % of all failures in PVC push fit joints. Arsénio et al., (2013) established that disconnection through joint rotation and axial pull out were the most important failure mechanisms for push fit joints (flexible joint) in PVC pipes. This is typically found where the angle of the pipe joint increases past 10° requiring a reduction in joint depth. Gasket failure is a result of age, drying out, loss of elasticity or degradation over time (Farrow et al., 2017). Kirby (1981) established that approximately 11 % of PVC failures were attributed to insufficient solvent and poor joint assemblage while 4 % arose from the use of excess solvent.

2.2.3 Pipe Coating and Lining

Pipe linings (inside the pipe) and coatings (outside the pipe) are used on pipes to slow the corrosion process (Mordak and Wheeler, 1988). Originally iron pipes were installed unprotected, but by the 1900s most pipes were dipped in bitumen to increase their service life. By the 1920s iron pipes were typically lined with cement mortar, which is more resistant to chemical degradation (Ruchti, 2017). Plastic pipes such as PVC and PE are not typically lined or coated due to their resistive properties to corrosion.

All modern DI and steel pipes are reliant on coatings and linings as their very thin walls (typically a few mm thick) are more susceptible to corrosion (Farrow et al., 2017). Linings include bitumen, cement mortar, synthetic resin and galvanisation, whilst coatings include resin or a PE sleeve (Trew et al., 1995). A multi-layered approach is now standard use for steel and DI, including polymer films (epoxy resin) or cement mortar linings, a bonded zinc and water-based paint coat or a mix of zinc and aluminium which acts like galvanisation but makes the surface active in preventing corrosion spreading. In environments that are highly corrosive to steel, additional protection can be included such as cathodic protection (Trew et al., 1995).

AC pipes are typically coated in bitumen, covered in a PE sleeve or coated in epoxy resin of a suitable grade (Farrow et al., 2017; Trew et al., 1995). Due to the natural resistance of AC to concrete corrosion, pipe coatings were only recommended when soils had pH values of less than 6.0, or sulphate between 0.8 – 2.0 % (Trew et al., 1995). AC pipes were not often lined, but when required three main types of material were used; firstly coal tar which was replaced by bitumen in the 1920s, and in later years, epoxy resin (Mordak and Wheeler, 1988). The performance of a pipe will be impacted by the type of lining or coating applied, so this information will be of interest to infrastructure modellers.

2.2.4 Manufacturing Defects

Defects introduced during manufacturing increase the probability of pipe failure. Defects in pit cast iron include: 1) non-uniform wall thickness, 2) porosity, 3) inclusions, 4) cold shuts and 5) micro-cracks (Trew et al., 1995) (Figure 2-5). Non-uniform wall thickness is a result of rising slag and off-centre inner core moulds, causing localised weakened areas of pipe wall. Porosity is caused by air trapped in the mould when the molten iron solidified, causing micro-crack formation paths. Inclusions are objects unintentionally introduced into the material fabric, breaking continuity, and causing subsequent weak points. Cold shuts are discontinuities in liquid streaming, resulting in incomplete fusion of two adjoining surfaces. Micro-cracks (longitudinal or transverse) are caused from uneven temperatures (i.e.

during cooling), forming weak points from which larger cracks formed under pressure (Makar et al., 2001; Rajkolhe and Khan, 2014; Thacker and Scholar, 2015).

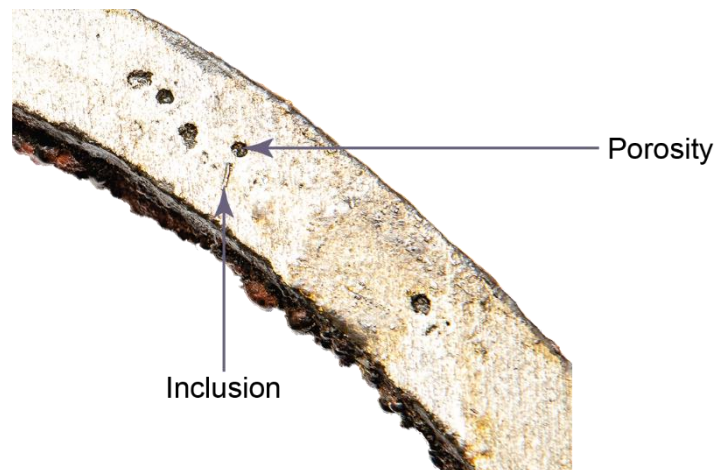


Figure 2-5 Porosity and inclusion defects in a cast iron pipe.

Spun iron, ductile iron and steel manufacturing processes create an even wall thickness, but common defects such as inclusions, porosity, cold shut and micro-cracks are still present. Steel also can develop defects from insufficient welding along the pipe tongue (Farrow et al., 2017). Uneven distribution of asbestos fibres in the cement matrix is the main defect in AC pipes. This results in pipe fracture where the pipe strength is weakened (Davis et al., 2008). PVC, like metal pipes, can have inclusions or porosity issues. Early PVC pipe (up to 1986), are considered to be low grade, often failing due to crack propagation from inclusions introduced during manufacturing. Poor gelation in early pipes (1970s -1980s) also resulted in low resistance against crack propagation. Breen, (2006) investigated PVC lifespan in the Netherlands and reported an optimal gelation range of 60-85 % of the pipe wall for high toughness. Some pipes produced in the 1970s had a gelation of < 40 % of the pipe wall whilst the 1980s was 50-60 %. For later PVC pipes with high levels of gelation (high grade) the main defects are micro-cracks present in pipes as a result of mechanical loading and environmental factors

during manufacturing (PVC can become brittle during manufacturing if the temperature is not high enough). Micro-cracks may result in slow crack propagation and eventual failure under cyclical pressure (Restrepo-Flórez, Bassi and Thompson, 2014; Awaja et al., 2016).

2.2.5 Pipe Damage from Handling, Storage and Third Parties

Poor pipe handling can cause invisible dents, cracks, and chips in protective coating. For example, DI is thin and light and easy to dent if moved with excessive force, and fragile coatings on many pipe materials are easily chipped (Farrow et al., 2017). Cracks formed due to careless handling have been observed in AC and plastic pipes but are not as typical as those on metal pipes. Installation issues also arise from poor storage, improper bedding, and poor pipe assemblage. Storage is a specific issue for PVC, as exposure to ultraviolet light for long periods of time embrittles the material. Improper bedding conditions, as reported by Al-Adeeb and Matti (1984), cause higher failure rates in AC pipes due to bending stresses from inadequate support. Kirby (1981) reported poor bedding as the single largest cause of failure in PVC pipes (13.7 % of all failures), where large stones caused point loading, and failures were largely due to a range of poor assemblage (joints, seal faults, over bending, excessive solvent etc.) accounting for 51.4 % of all failures in the observed UK network. Ruchti (2017) identified third party construction as an issue, where excavation within the vicinity of buried water pipes is a significant risk, and scraping, moving, or unsettling bedding areas can result in eventual failure. Scraping of PE pipes > 10 % of the wall thickness can also lead to premature failure.

2.2.6 Corrosion and Chemical Degradation

Corrosion and chemical degradation deteriorates pipes and affects their material integrity through pipe wall thinning, and functional integrity through pipe displacement affecting joints (Vreeburg et al., 2013). Chemical degradation of metal pipes is well documented (Babovic et al., 2002; Folkman, 2018; Hou et al., 2016; Makar et al., 2001; Rajani and Tesfamariam, 2004; Wasim et al., 2018)

and rates of failures vary between geographic region. For example Folkman's (2018) US study found that 28 % of all pipe failures were caused by corrosion, with iron pipes having the highest failure rate, whilst Babovic et al., (2002) found 48 % of iron pipe failures in a study in Denmark were a result of corrosion. Whilst the literature may vary, it is accepted that corrosion of metal pipes has a significant impact on pipe failure. The mechanism for corrosion depends on soil properties, which differ for each location (Hou et al., 2016), thus highlighting the importance of using localised soil maps and data (Pritchard et al., 2013). Wasim et al., (2018) provides a comprehensive review of literature on soil corrosion on pipes. In summary, corrosion rates are related to the coupled effects of a number of soil parameters, which broadly results in higher corrosion rates when: pH is low, moisture content increases (until a limit is reached and then they decline), soil resistivity is low < 2000 ohm cms, soils are highly aerated (but only when soil moisture is present), high temperatures and high levels of soluble salts are present. The interaction between soil parameters is complex and will manifest in differing outcomes for each material. Wasim et al., (2018) reported that iron corroded faster than DI and steel, and failures in iron pipes can occur suddenly and result in catastrophic failure, whilst DI and steel typically fail in the form of leaks.

Rezaei, Ryan and Stoianov (2015) reported that many failures occurred in aggressive soils where corrosion initiates and accelerates pipe failure. Karpachevskii, Goroshevskii and Zubkova (2011) found that disturbed bedding and backfill can increase corrosion, where soil aeration and water permeability increased this process. Hu, Wang and Chowdhury (2013) found acidic soils with high concentrations of sulphate increase corrosion rates.

Highly corrosive soils can deteriorate pipes at faster rates, especially for older pipes without protection. Two main types of corrosion occur in metal pipes: 1) graphitization and 2) corrosion pitting (Figure 2-6). Graphitisation, described by Makar, Desnoyers and McDonald (2001), occurs mainly in cast iron, and is a key corrosion process that occurs when iron oxide reacts with the graphite in the iron

alloy, leaching the iron and leaving the graphite behind. In some cases of aggressive graphitisation all iron is removed, significantly weakening the pipe. (Farrow et al., 2017; Rajani and Kleiner, 2001). Corrosion pitting occurs in all forms of metal pipes, from the presence of sulphate and chloride ions in the surrounding environment, causing deterioration of the pipe material from electrochemical and or chemical action corrosion (Volk et al., 2000).



Figure 2-6 Cast iron pipe corrosion and chemical degradation.

Two main types of concrete corrosion occur in AC pipes: 1) lime leaching and 2) sulphate attack (Hu et al., 2013). Early AC pipes were manufactured using free lime, which was susceptible to lime leaching in certain soil conditions, being particularly evident when silica was not used (Punurai and Davis, 2017). Lime leaching occurs when external pipe conditions are acidic with a low ion content mobile water source (typically a high groundwater level) (Hu et al., 2013; Silva et al., 2002). In such an environment, calcium hydroxide (a free by-product from manufacturing cement mortar, used to stabilise hydrated silicates) is dissolved out of the cement matrix and into the groundwater (Rozière et al., 2009). Once the calcium hydroxide has leached from the cement matrix, the hydrated silicates in the cement mortar decompose. This results in hole formation within the pipe

matrix, eventually leading to a soft material where the cement no longer binds the asbestos (Gong et al., 2016). Lime leaching in pipes can also occur when conveying very soft water ($\text{pH} > 7$ and a hardness of $< 10 \text{ mg/l}$ of calcium and magnesium carbonate) which promotes lime leaching and eventually leads to pipe failure (Al-Adeeb and Matti, 1984). Sulphate attack occurs when sulphate in the surrounding groundwater reacts with the calcium hydroxide in the AC material to form calcium sulphate. The calcium sulphate reacts further in the AC matrix and expands, cracking and breaking the cement. AC failure rates were found to be accelerated in soils containing aggressive ions such as magnesium, chloride and sulphate and where pH levels are less than 6.3 (Hu et al., 2013).

VC and PE are both resilient to corrosion (Ellison and Spencer, 2016). However, both can deteriorate under the right conditions. Both PE and PVC are vulnerable to organic chemicals from polluted soils and from chlorinated water (Kowalska et al., 2016). Soil pollutants can diffuse through the pipe affecting the quality of drinking water (Holder et al., 2019), whilst chlorinated water as a result of using disinfectants (such as chlorine or hypochlorous acid), deteriorates the wall of the pipe through oxidation, which over time embrittles the material, resulting in crack propagation driven by internal pressure. Oxidative degradation can also occur as a result of spontaneous chemical reaction with atmospheric oxygen (Colin et al., 2009; Ghabeche et al., 2015; Mikdam et al., 2017). PVC can also deteriorate due to abiotic factors (e.g., cold temperatures or exposure to UV) and biotic factors (e.g., soil microorganisms or roots) which can result in brittle fracture as the pipe weakens. Brittle fracture in PVC was reportedly exacerbated in contaminated soil and with internal water with high levels of '*active organic compounds (e.g. detergents or solvents)*' are present (Trew et al., 1995, p. 257), and after long periods of exposure to UV (Ruchti, 2017).

2.2.7 Pipe Age

While pipe age has been reported to have a linear relationship with failure rates, with older pipes being more likely to fail (Wengström, 1993), the relationship is actually more complex. Failure rates can be expected to be higher in the months

immediately following installation, then dropping to a low rate of failure for a number of decades, before increasing with age. Folkman (2018) collected data from 308 water companies in the USA and Canada and reported that the average age of pipe failure is 50 years. Age can also correlate with periods of uniform manufacturing, installation and operational practices (Andreou et al., 1987a; Kettler and Goulter, 1985; Wengström, 1993). Figure 2-7 shows the relationship between failure rates and installation period for the different pipe materials present in the utility providers network.

Failure rates for iron, DI and steel peak for pipes installed during the 1920s – 1940s, AC during the 1940s – 1960s. Both PE and PVC have steady failure rates by age. Iron shows that failure rates increase for newer pipes, up to the 1920s – 40s and then fall again which is unexpected. This could be due to manufacturing, where materials and standards have been manufactured differently, resulting in different quality of materials (Wols and Van Thienen, 2016); early iron pipes were locally cast and had very thick walls. Iron pipe standards were then introduced in 1917 with BS 78 1917, where pipes were mainly spun to a specification which produced thinner walls that failed earlier than thicker pipes. Thereafter failure rates fall up to the 1960s, which is due to the shift towards centrifugally spun iron and pipe protection, constructional methods which produced a more resilient pipe or the replacement of these lower quality pipes with newer materials (Wols and Van Thienen, 2016). All material failure rates start to improve after the 1940s – 1960s, perhaps due to the age of the installation, improved construction practices and improved protection against corrosion.

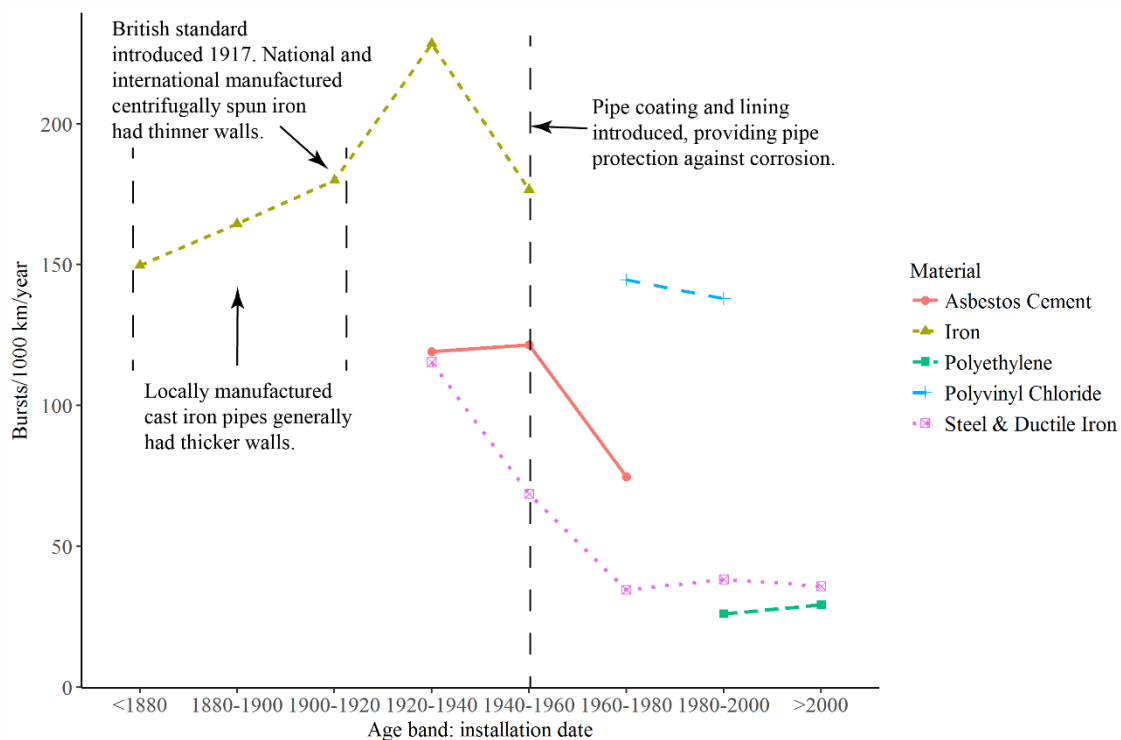


Figure 2-7 Summary of introduction year and successive failure rates by age by material type (Utility Provider, 2018).

2.2.8 Pipe Diameter

Pipe failure records suggest a strong relationship between failure rates and pipe diameter size (Hu and Hubble, 2007; Kettler and Goulter, 1985; Kimutai et al., 2015; Pelletier et al., 2003). The highest failure rate was reported in pipes < 200 mm in diameter, however, it has regularly been noted that this diameter range is also the most frequent (Bruaset and Sægrov, 2018; Fuchs-Hanusch et al., 2013; Gould et al., 2011; Hu and Hubble, 2007). Higher failure rates in small diameter pipes may be associated with low resilience to ground movement and corrosion (thinner walls), poor joint reliability (Gould et al., 2013) and susceptibility to nearby construction activities especially in urban areas and at shallower depths (Bruaset and Sægrov, 2018). Circumferential failures are typically associated with pipes < 200 mm in diameter, and are the most frequent failure mode for metal and AC pipes of this diameter (Bruaset and Sægrov, 2018; Wengström, 1993). Where longitudinal failures occurred in smaller diameter pipes, Ruchti (2017) reported

differential settlement as the common mechanism for failure. When researching large pipes Rajeev et al. (2014) found longitudinal failures and holes to be the common failure mode for larger pipes, typically > 300 mm in diameter; the evaluation suggesting corrosion and internal water pressure as the main causes. Typical failure rates by diameter size and material type are presented in Figure 2-8.

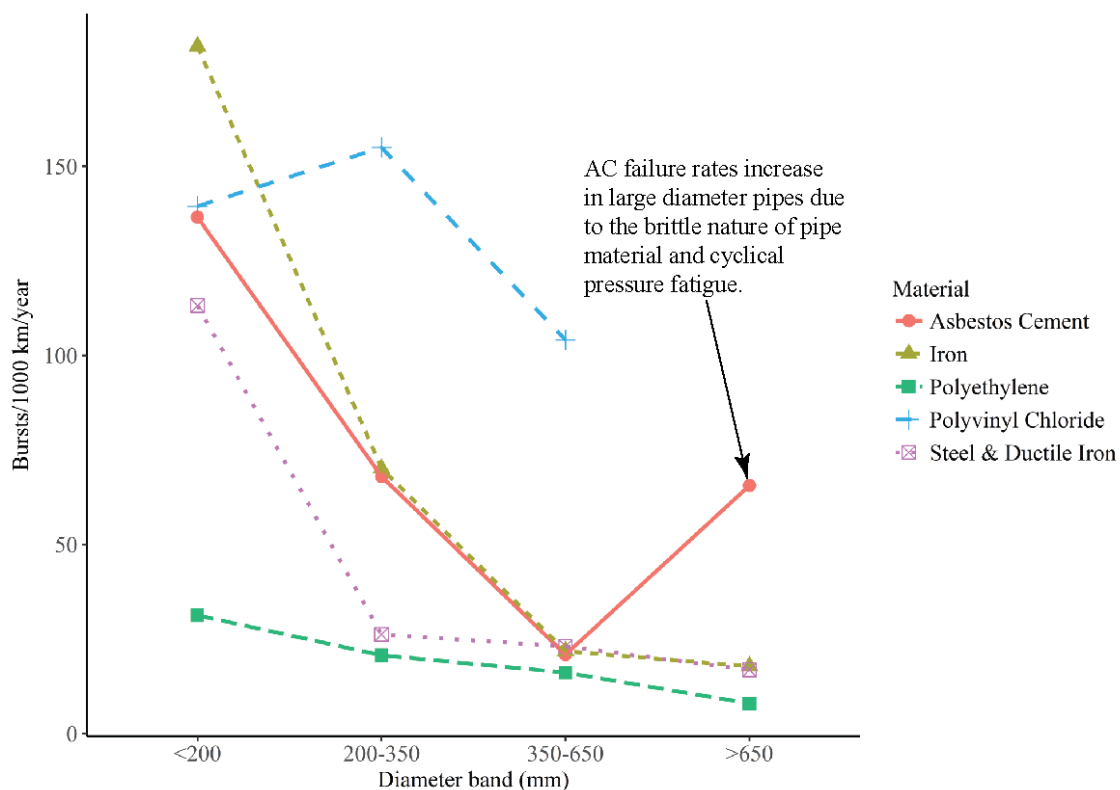


Figure 2-8 Summary of failure rates by diameter size and pipe material type (Utility Provider, 2018).

2.3 Environmental Factors Influencing Pipe Failure

It is apparent that pipe intrinsic factors have a strong influence on the performance of the pipe. Nevertheless, pipes installed in non-aggressive soils with stable weather conditions will fail less often than those installed in more aggressive soils with highly fluctuating temperatures and levels of precipitation.

Thus, predictive models of pipe failure will also perform better with the addition of data on preceding and localised environmental conditions.

2.3.1 Seasonality

Sudden or dramatic seasonal changes have a strong correlation with failure rates (Fuchs-Hanusch et al., 2013; Gould et al., 2011; Laucelli et al., 2014; Pritchard et al., 2013; Wols and Van Thienen, 2014). Failure rates are higher during dry summers and autumns (typically for AC and PVC pipes) or cold winters (iron, DI and steel), the opposite was observed for wet summers or mild autumn and winters. Fuchs-Hanusch et al., (2013) reported that failure frequency was higher in rigid pipes up to 200 mm during the winter. Andreou (1986) found that circumferential failure is more prevalent in pipes up to 150 mm during the summer. Kottman (1988) reported that season has little influence on pipes at greater depth (> 1 m) and on larger pipes, since they generally fail under high pressure and not ground movement resulting from seasonal weather changes.

A comparison of seasonal failure from the utility provider is shown in Figure 2-9. The results show that highest failure rates for iron are during the winter months, whilst AC and, to a lesser extent, PVC have higher failure rates during the summer. DI, steel, and PE show little variation to seasonal change but can be seen to slightly increase during the winter.

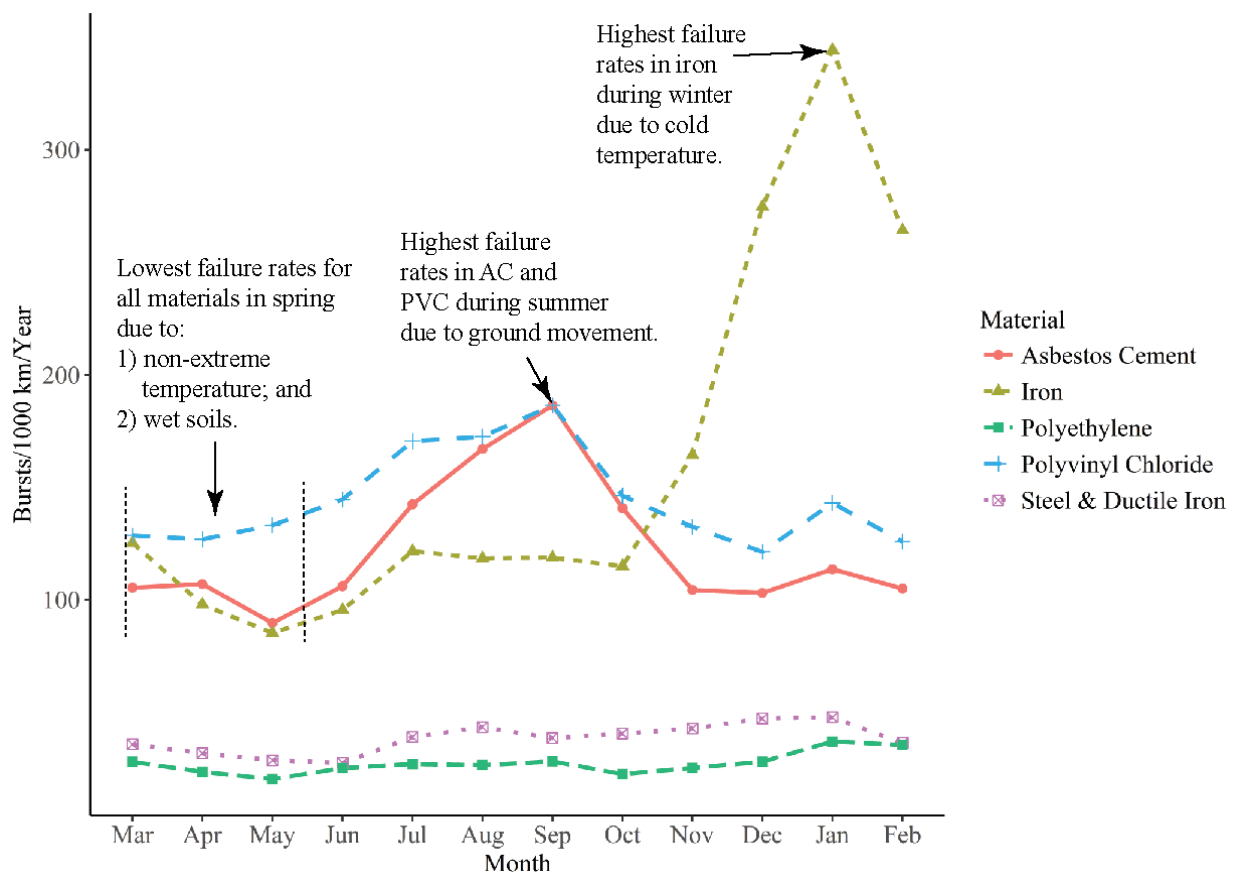


Figure 2-9 Summary of seasonal failure rates by material type (Utility Provider, 2018).

After reviewing a number of studies considering how weather patterns influence seasonal variations in pipe failure rates, Wols and Van Thienen, (2014) established the most influential weather factors to be temperature, frost and rainfall deficit (a surrogate for soil moisture deficit (SMD) (Laucelli et al., 2014)), although the effects will vary by pipe material and geographical region (Gould et al., 2011).

2.3.2 Cold Temperature

Seasonal pipe failure as a result of temperature has been well documented (Bruaset and Sægrov, 2018; Clayton et al., 2010; Gould et al., 2011; Fuchs-Hanusch et al., 2013; Hu and Hubble, 2007; Rajani and Kleiner, 2001), concluding that most failures occur during winter months (approximately 60 %

(Rezaei et al., 2015)). Wols et al., (2019) observed the effects of weather on pipe failures in the Netherlands, and found that iron, DI, PVC, and AC pipes have a varying correlation with cold temperatures. Iron is the most susceptible to cold temperatures below 3°C, and winter failures may be a direct response to the effects of freezing and expansion of soil moisture causing tensile forces and resulting in ground movement and additional compression on the pipe resulting in circumferential pipe failure (Babovic et al., 2002; Kakoudakis et al., 2018; Balvant Rajani et al., 1996). Studies in Norway (Bruaset and Sægrov, 2018) and the US and Canada (Folkman, 2018) show that temperatures reach low enough to cause frost heave, which Rajani et al., (1996) report result in movement of trench and side fill. However, in the UK with moderate climates, soil displacement is more likely to occur in the upper soil layer during freezing and thawing (Wols and Van Thienen, 2014). The highest failure rates found during winter are associated with the first winter frost, where weaker pipes fail, leaving a more resilient network for the remaining winter frosts (Newport, 1981). Subsequent reductions in peak low temperature following the first frost also observes higher failure rates (Habibian, 1994).

The effects of frost are typically seen at depths shallower than 0.5 - 1.0 m below the soil surface depending on frost duration (Pritchard et al., 2013). As guidelines require pipe installation 0.75 m and 1.35 m below the soil surface (Anglian Water, 2014), Habibian (1994) suggests that frost penetration to the depths of buried pipes requires a sustained period of frost. This was also supported by Newport, (1981) who observed failure data from Severn-Trent Water plc. and revealed failure rates for iron increased with deeper frost penetration. Bruaset and Sægrov, (2018) also reported that poor bedding conditions can exacerbate the effects of frost heave. Summers with long periods of dry weather lead to soil with a low latent heat capacity, which results in deeper frost penetration into the soil during winter (Hu and Hubble, 2007; Laucelli et al., 2014).

Pipe failure due to large temperature differences (Pietrucha-Urbanik, 2015) were observed, or where the temperature dropped and remained low for a long period

(a cold snap) (Edil and Bahmanyar, 1983). Farrow et al., (2017) suggest that thermal stress from cold reservoir water can contract pipes, resulting in consequent circumferential fractures, and that this has a higher influence on failure rate during the winter than frost. Rajani, Kleiner and Sink (2012) further observed an increase in failure rate during fast temperature transits for metal pipes below 0°C, suggesting pipes have lower resistance to rapid changes in ground and conveyed water temperature. Rajani and Tesfamariam (2004) reported that where there was a significant temperature difference between internal water (1-2°C) and the external proximal soil (10-12°C), there was a consequent increase in failure rate.

The ability of plastic pipes (PVC and PE) to withstand thermal expansion and contraction (plasticity) means that in general, temperature variation and winter frost has little influence on failure rates (Ruchti, 2017) (Figure 2-10). However, Wols and Van Thienen, (2014) reported that PVC pipes have high joint failures during cold temperatures. High joint failure was also reported by Wols and Van Thienen, (2014) for AC pipes, whilst Hu and Hubble (2007) reported that AC failure rates increased during the winter where long periods of consecutive days' air frost below 0°C were observed. Plastic pipes do not generally fail through frost loading due to their flexible nature, it has been recorded in Canada and the US that pipes can fail where low temperatures cause the internal water to freeze and expand, placing pressure on the pipe (especially where defects are present) (Edwards et al., 2009) – freezing water can expand by 9 % in volume. Furthermore, prolonged frost periods can cause increased brittleness in plastic pipes resulting in premature failure (Rezaei et al., 2015). Future trends in climate change suggests that milder temperatures during winter will result in fewer pipe failures during this period, especially in iron pipes (Wols and Van Thienen, 2016).

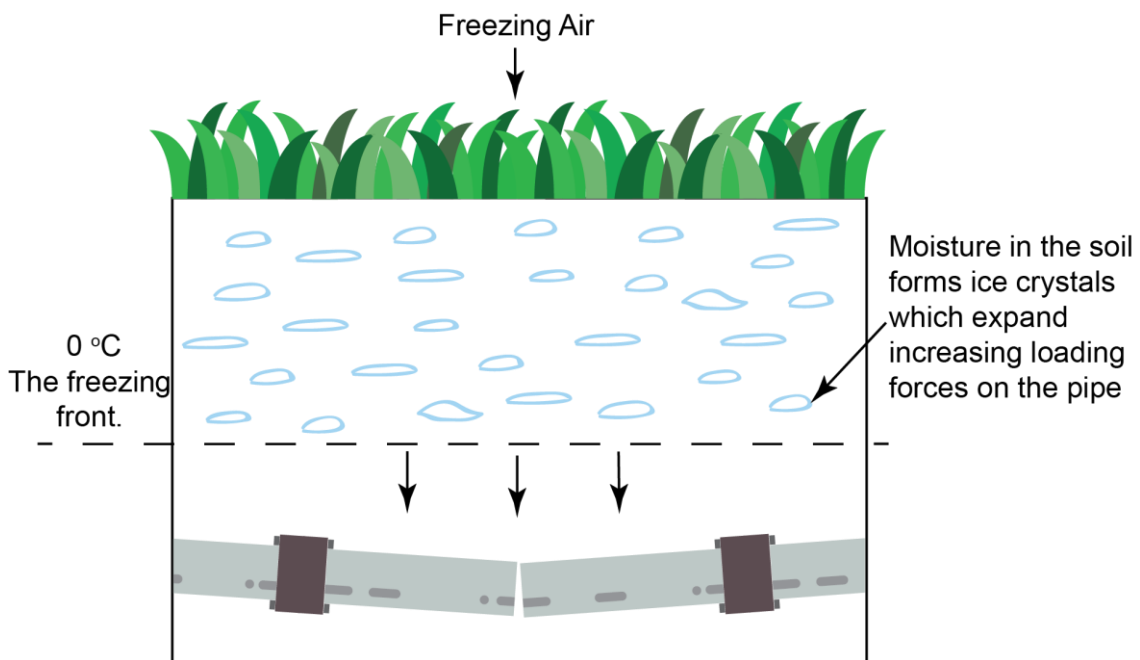


Figure 2-10 Frost loading on pipe.

2.3.3 Warm Temperature and Soil Movement

Soil Moisture Deficit (SMD) represents the difference between soil moisture content and soil moisture saturation. SMD within clay and peat rich soils can lead to shrinkage and swelling resulting in differential ground movement and multi-directional pressure on the pipe resulting in failure (Gould et al., 2011). Farewell, Hallett and Truckell (2012) highlighted the effects of SMD on buried assets, identifying that failure rates are higher in the summer (particularly AC) when SMD is high. Similar findings were reported by Wols and Van Thienen (2014) but using temperature and rainfall deficit to find drought conditions. The results showed that AC failures increased in the summer months as a result of differential ground movement. Mordak and Wheeler (1988) highlighted that most failures observed for AC during the summer were circumferential fractures, a common failure mode from bending stresses on the pipe or joint disconnection, and Hu, Wang and Chowdhury (2013) observed the same, typically for pipes between 100 - 150 mm in diameter. Alternative suggestions for failures also included thermal stress on the pipe as a result of the difference in temperature between the soil and internal

water or longitudinal expansion the pipe (Balvant Rajani et al., 1996). Wols and Van Thienen (2014) found an increase in steel pipe failure during the summer, suggesting thermal expansion at high temperatures on pipes with a reduced resilience (due to degradation) as a potential cause.

Pipe failure is lowest during the spring when soils are consistently wet, so movement is unlikely. Chan, Gould and Davis (2005) suggested that the late summer and early autumn months are the most influential time of year for AC failure rates, since soil moisture and temperature varies the most during this period, increasing the shrink swell effect. This is also shown in the data from the UK utility provider in Figure 2-10. UK soils normally dry to 1 – 1.5 m (O. G. Pritchard et al., 2015a), but long periods of drought can result in deeper drying and higher failure rates (Farewell et al., 2012b). High failure rate in PVC push fit joints are associated with ground movement causing disconnection through joint rotation and axial pull (Arsénio et al., 2013). The effects of soil shrink swell on pipes from changes in soil moisture are shown in Figure 2-11.

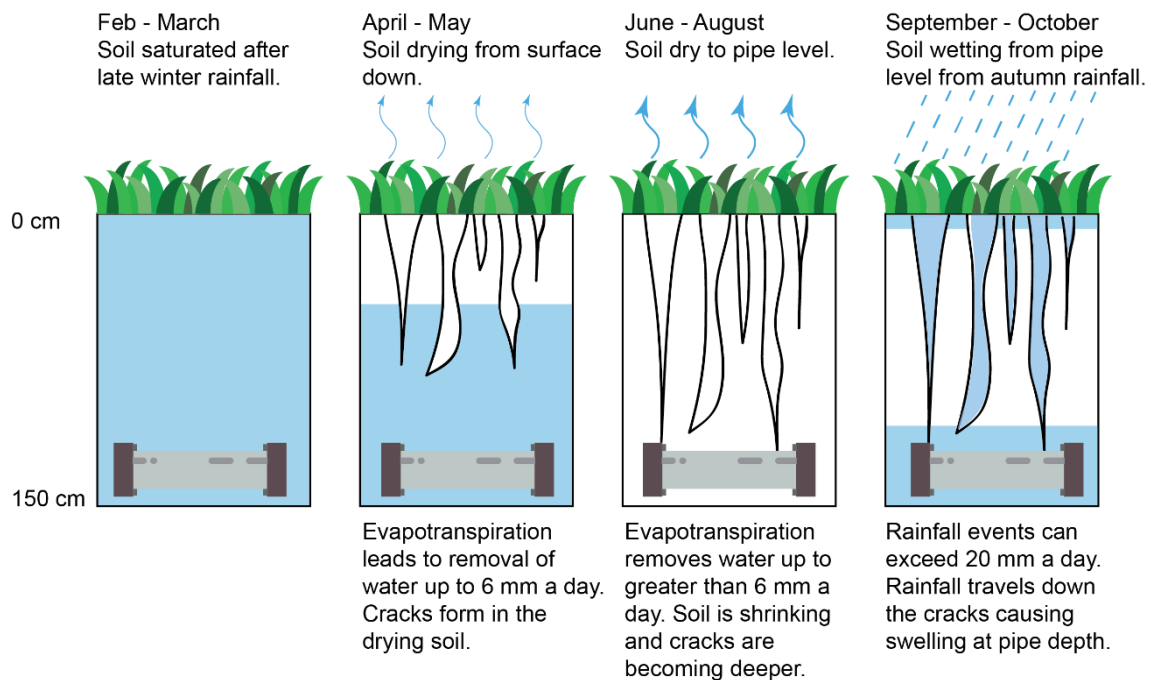


Figure 2-11 Seasonal shrink swell effect in clay-rich soils (after Farewell et al., 2012, p. 10).

Soil related ground movement represents a hazard to pipes, and depends on soil properties (i.e., texture, structure, porosity) which control moisture transport and storage. Soils are generally considered to be weaker when wetter, and are therefore more prone to movement (Pritchard et al., 2013; Yahaya et al., 2011). Clay and peat soils have fine particles, low permeability and poor aeration which result in high water retention after wet periods (swelling) and high evapotranspiration when temperature is high and rainfall is low, leads to soil shrinkage. Clay and peat soil shrink swell is associated with circumferential breaks and joint failure in pipes, and are typically associated with soils which cause higher rates of chemical degradation to pipes (Hu and Hubble, 2007; Mordak and Wheeler, 1988; Pritchard et al., 2014; Ruchti, 2017).

Patterns of future climate change suggest a bearing on geographical variation in environmental vulnerability to soil-related influences, where prolonged droughts cause increased soil movement (O. G. Pritchard et al., 2015b; Pritchard et al.,

2014), and potentially higher failure rates during this period, especially in AC pipes (Wols and Van Thienen, 2016).

2.3.4 Other Soil Ground Movement Hazards

Other soil related ground movements include peat shrinkage, sand washout, silt ground heave and soft and compressible soils (Pritchard et al., 2013). Sandy soils typically drain water due to their large particle size. Sandy soils that contain particles between 0.06 mm to 2.0 mm are susceptible to wash out and erosion from excessive water flow (i.e. events such as pipe failure) resulting in pipes being unsupported and sagging (Brink et al., 1982). Farewell, Jude and Pritchard, (2018) reported that pipe failures in sandy soil had a spatiotemporal relationship with previous failures, suggesting loss of water under pressure can lead to washout of sandy soils and further failures. They also reported that trenching during construction of roads provides preferential hydrological pathways which can lead to washout cavities in sandy soil conditions. Soft soils such as organic peat (organic soils) have poor bearing capacity due to their generally high retention of moisture, which offers little support to pipes. The result is a sagging, or 'bridged', pipe with pressure loading causing failure at the pressure point (O. G. Pritchard et al., 2015b).

Differential settlement is the natural compaction of soils which results in the uneven settlement of the pipe. This can be caused by natural soil transition, poor foundation construction or crossing with other utilities (Babovic et al., 2002). The transition of pipes across different soils can result in premature pipe failure, where pipes curve as a result of soft to hard soil transition, especially more rigid pipe materials (Wols et al., 2014). Wols and Van Thienen, (2014) suggests that climate change and potential increase in long periods of drought will increase the effects of differential settlement on pipes.

Once the pipe intrinsic and environment factors are sufficiently considered, the way in which the pipes are operated, managed, and used should be investigated.

In the authors' experience, it is this suite of operational factors for which it is the most difficult to obtain consistent reliable data over sustained periods of time.

2.4 Operational Factors Influencing Pipe Failure

While it is recognised in the industry that pipe management and operation is linked to the rate of failure, there is less discussion in the literature on these impacts specifically discussing failure rates. This may be because operational data is less commonly recorded in model-friendly data sets. We discuss common operating factors but recognise the following as potentially influential to pipe failure: pressure management, flushing events, network changes such as non-return valve replacement, jetting and maintenance, network jobs, fire hydrant testing, water temperature and proximity to high water consumers (e.g., agricultural facilities).

2.4.1 Internal Water Pressure

Changes in internal pressure can increase the likelihood of pipe failure. Two main influences in changes to internal pressure have been reported to cause early pipe failure: 1) cyclical pressure and 2) transient surge pressure. Cyclical operational pressure from consumer patterns and changes in network operations (pressure management) leads to periods of high and low pressure. Rezaei, Ryan and Stoianov (2015) studied the effects of cyclical pressure on a water supply network and found that cyclic loading can result in fatigue failure if the frequency and magnitude were appropriate. This was especially apparent in pipes with defects where the loading stress can speed up crack propagation resulting in early failure. Wols et al., (2019) reported an increase in AC failure rates during high temperatures and suggested water demand causing high internal pressure or a large pressure difference could be a possible cause. Ilić, (2009) reported for a single district meter area in Zagreb that pressure regulation had a positive effect on reducing pipeline failures by 17 %. This was particularly evident in iron and PVC pipes, since joints in these materials have less resilience to pressure oscillations.

Transient surge pressure also known as “water hammer” is caused by a result of network operations such as flushing events, fire hydrant testing, valve changes or pumping station failure. These can cause additional stress on pipes through sudden changes in internal pressure exposing the pipe to the strongest physical load it is likely to experience (Martínez-Codina et al., 2016). Pozos-Estrada et al., (2016) researched pressure on concrete pipes and found that transient pressure over pressurizes the pipe causing catastrophic ruptures of the pipe. Martínez-Codina et al., (2016) reported on a range of pipe materials including metal, cement, and plastic, and suggested that medium to high pressure inside the pipe caused during transient surge pressure is one of the common causes of pipe failure in large diameter pipes and is especially prevalent in pipes where micro-cracks and corrosion have weakened their resilience. The result is loss of integrity or pipe wall through blow outs or joint failure (Boulos et al., 2005; Rezaei, Ryan and Stoianov, 2015). Further to this, it has been reported that the presence of air pockets in pressurised pipelines can greatly exacerbate the effects of transient pressure (Pozos-Estrada et al., 2016).

Boulos et al. (2005) reported transient surge pressure as being less likely to affect urban distribution networks due to short pipe length (less than 600 m), suggesting short length of pipes tend to limit the effects of transient surge pressure due to junctions causing pressure wave reflections which dissipate the effects of concussive water hammer. Operation of pumping stations can result in failures where power cuts can cause reverse flow surge, and transient surge pressure introduced after re-joining the pump (Guyer, 2013). Ruchti (2017) highlighted that smaller diameter pipes are less prone to failure than larger pipes from internal pressure, largely due to the low pressures commonly used in such pipes.

2.4.2 Previous Failures

Previous failures can result in cascading failures where the initial failure can result in a secondary or multiple consecutive failures located spatiotemporally close to the first (Clark, Stafford and Goodrich, 1982). Many statistical models do not account for this phenomenon despite previous studies suggesting its importance

(Scheidegger et al., 2015). Goulter and Kazemi (1988) revealed that 22 % of failures occurred within one meter of the previous failure and 42 % of these failures occurred within 1 day of the first failure. This pattern is typically attributed to the aging deteriorating pipe and the potential for the first failure to disturb the surrounding environment through washing and eroding bedding conditions, altered soil moisture conditions weakening soil and the disturbance of the pipe during repair operations (Farewell et al., 2018; Farrow et al., 2017).

2.5 Discussion

The failure modes and mechanisms of water distribution pipes are complex, varying across geographical locations and by pipe material. Factors that increase the chance of failure are often interconnected, and thus typically no single failure mechanism is uniquely related to one factor. However, certain factors emerge as having a greater influence than others for each failure mode. Table 2-3 shows a summary of the failure modes and mechanisms by material type.

Table 2-3 Summary of the failure modes and mechanisms by material type.

Material	Mechanism			Typical Failure Mode
	Pipe - Intrinsic	Environmental	Operational	
Iron	Pipe diameter. Manufacturing defects. Graphitization. Pipe protection. Rigid joints. Construction and repair – accidental damage.	Cold temperatures. Frost. Cold internal water temperature. Highly corrosive soils.	Cyclical pressure fatigue. Transient pressure. Management operations.	Circumferential break. Joint failure. Longitudinal failure. Chemical attack.
Steel and Ductile Iron	Thin pipe wall. Manufacturing defects.	Highly corrosive soils.	High pressure.	Chemical attack. Joint failure.

Material	Mechanism			Typical Failure Mode
	Pipe - Intrinsic	Environmental	Operational	
	Pipe protection. Rigid joints.		Cyclical pressure fatigue.	
Asbestos Cement	Pipe diameter. Manufacturing defects. Pipe protection. Rigid joints.	Warm temperatures. Low rainfall. Fluctuating soil moisture. Clay and peat soils – shrink swell potential. Highly corrosive soils.	High pressure. Cyclical pressure fatigue.	Circumferential break. Joint failure. Chemical attack. Longitudinal failure.
Polyvinyl Chloride	Poor joint assemblage – solvent. Storage – UV light exposure Manufacturing defects. Loading – sensitivity to point loads.	Warm Temperatures. Low rainfall. Fluctuating Soil Moisture. Clay and peat soils – shrink swell potential. Sandy soils – wash out	High pressure. Cyclical pressure fatigue.	Joint failure. Longitudinal failure.
Polyethylene	Poor joint assemblage.	-	-	Joint failure. Longitudinal failure.

Iron pipes have a high failure rate during the winter as a result of low temperatures and frost, which results in freezing soil moisture and expansion of soils, causing circumferential failures or joint failure (especially in rigid joint systems). This results in failure in the less resilient pipes (those affected by highly corrosive soils). Corrosion is a major factor in iron pipe failure, particularly before coating was introduced or where the coating was damaged during installation. Corrosion from soils causes thinned pipe walls which results in pin holes and small leaks in small diameter pipes, to blowout holes in larger pipes if high pressure is reached.

Longitudinal failures in iron pipes are less common, but typically, result from defects in the manufacturing process which leads to micro crack propagation and eventual failure. As corrosive soils also tend to be shrinkable, the ground movement associated with summer and autumn can also break weakened pipes.

Steel and Ductile Iron have a low failure rate but are prone to corrosion holes in pipes and potential blowout holes in large pipes with high pressure. Rapid corrosion in the thin walls can occur when the protective coating is damaged. Seasonal weather variation does not appear to strongly impact the failure rates of these pipe materials.

AC pipes largely fail during the summer months as a result of differential ground movement. Ground movement causes multi-dimensional stress on pipes which lead to eventual failure of weakened pipes, especially those with manufacturing defects and pipes which have been deteriorated by highly corrosive soils. The main factors affecting AC pipes include seasonally fluctuating temperatures, and rainfall leading to large changes in soil moisture, and associated shrink-swell, in clay rich soils. Circumferential failure is the most common mode of failure accounting for approximately 75 % of all failures, being typically associated with pipes < 200 mm in diameter due to thin walls and low resilience to bending especially when weakened by chemical attack.

PVC pipes have a higher failure rate during the summer and the majority of failures occur from joint failure. This is a result of differential ground movement which results in joint disconnection through joint rotation and axial pull out in push-fit joints and poor assemblage. Poor bedding conditions, such as coarse rock, may cause pressure points on PVC pipes and can lead to longitudinal failure starting from the pressure point. Cyclical pressure fatigue can cause longitudinal failures in pipes with manufacturing defects, where micro cracks propagate quickly, or where the pipe becomes brittle from the effects of UV light during storage.

PE pipes have very low failure rates and are resistant to corrosive soils, ground movement, weather conditions and pressure changes from operational management. For these reasons, they are widely used in new development and asset upgrade. Most failures on PE pipes come from poor construction of joints since specialist training is required for the electrofusion fitting.

There exists a wealth of information in existing literature that can be used by data scientists to explore pipe failures using statistical models. The information presented here can be used to guide initial exploratory investigations to ensure that relevant data inputs and modelled outputs are consistent with existing understandings of typical modes and mechanisms for pipe failure. Operational factors are very important, but data is still inconsistent and incomplete across large networks (e.g., pressure is still largely modelled, rather than measured across much of the UK water network). As this situation improves, it is highly likely that recording operational data across entire water networks in a consistent manner over many years will give rise to improved predictive models of pipe performance in the future. Modelling of the impact of management decisions, investments and public engagement schemes will become an important area that requires further investigation.

2.6 Conclusion

There are a wide variety of factors which increase the likelihood of pipe failure. Pipe failures are more likely on older less resilient pipes, in aggressive soils where weather and operational conditions are extreme and unstable. Sudden changes in temperature, pressure or soil moisture levels will increase internal and external stresses on pipes, increasing the chance of failure.

When building data-driven, statistical models of pipe failure, it is important to obtain a basic understanding of the common mechanisms of pipe failure. When this step is overlooked, data can be misused through misunderstanding (especially common is the confusion of correlation and causation). Sense-checking models, and their data inputs, against modes and mechanisms of failure

reported in the literature will help ensure that data driven models of pipe failure are robust and fit for purpose. Such models, in turn, can lead to actionable information which utilities can use to improve the performance of this critical national infrastructure and meet the water demands of a growing population under a more extreme climate.

2.7 Author Contributions

Neal Barton: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing & Original Draft, Visualization, Data Curation. Timothy Farewell: Project Administration, Conceptualization, Validation, Formal Analysis, Resources, Writing, Review & Editing, Supervision, Data Curation, Funding Acquisition. Stephen Hallett: Writing, Review & Editing, Supervision, Funding Acquisition. Tim Acland: Writing, Review & Editing.

2.8 References

- Al-Adeeb, A.M., Matti, M.A., 1984. Leaching corrosion of asbestos cement pipes. *Int. J. Cem. Compos. Light. Concr.* 6, 233–240. [https://doi.org/10.1016/0262-5075\(84\)90018-6](https://doi.org/10.1016/0262-5075(84)90018-6)
- Andreou, S., 1986. Predictive models for pipe break failures and their implications on maintenance planning strategies for deteriorating water distribution systems. PhD Thesis.
- Andreou, S.A., Marks, D.H., Clark, R.M., 1987. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. *Adv. Water Resoure.* 10, 2–10. [https://doi.org/10.1016/0309-1708\(87\)90002-9](https://doi.org/10.1016/0309-1708(87)90002-9)
- Anglian Water, 2018. Drinking water network failure images.
- Anglian Water, 2014. Laying of Water Supply Pipes Prior To Connection, Anglian Water.
- Arsénio, A.M., Pieterse, I., Vreeburg, J., de Bont, R., Rietveld, L., 2013. Failure

- mechanisms and condition assessment of PVC push-fit joints in drinking water networks. *J. Water Supply Res. Technol.* 62, 78–85. <https://doi.org/10.2166/aqua.2013.026>
- Asnaashari, A., McBean, E.A., Gharabaghi, B., Tutt, D., 2013. Forecasting watermain failure using artificial neural network modelling. *Can. Water Resour. J.* 38, 24–33. <https://doi.org/10.1080/07011784.2013.774153>
- Awaja, F., Zhang, S., Tripathi, M., Nikiforov, A., Pugno, N., 2016. Cracks, microcracks and fracture in polymer structures: Formation, detection, autonomic repair. *Prog. Mater. Sci.* 83, 536–573. <https://doi.org/10.1016/j.pmatsci.2016.07.007>
- Babovic, V., Drecourt, J.-P., Keijzer, M., Hansen, P.F., 2002. A data mining approach to modelling of water supply assets. *Urban Water* 401–414.
- Beuken, R.H.S., Mesman, G.A.M., de Kater, H., 2012. The Application of In-Line Inspection Technology for Condition Assessment of Water Mains 41203, 991–1001. [https://doi.org/10.1061/41203\(425\)90](https://doi.org/10.1061/41203(425)90)
- Boulos, P.F., Karney, B.W., Wood, D.J., Lingireddy, S., 2005. Hydraulic transient guidelines for protecting water distribution systems. *J. / Am. Water Work. Assoc.* 97, 111–124. <https://doi.org/10.1002/j.1551-8833.2005.tb10892.x>
- Breen, J., 2006. Expected lifetime of existing PVC water distribution systems management summary. The Hague.
- Brink, A.B.A., Partridge, T.C., Williams, A.A.B., 1982. Soil survey for engineering, Monographs on soil survey. Oxford University Press., Oxford.
- Bruaset, S., Sægrov, S., 2018. An analysis of the potential impact of climate change on the structural reliability of drinking water pipes in cold climate regions. *Water (Switzerland)* 10. <https://doi.org/10.3390/w10040411>
- Burn, S., Davis, P., Schiller, T.L., Tiganis, B.E., Tjandraatmadja, G.F., Cardy, M., Gould, S., Sadler, P.A., Whittle, A.J., 2005. Long-term performance

prediction for PVC pipes.

Chan, D., Gould, S., Davis, P., 2005. Data analysis and laboratory investigation of the behaviour of pipes buried in reactive clay 206–211.

Clark, R.M., Stafford, C.L., Goodrich, J.A., 1982. Water distribution systems: a spatial and cost evaluation. *J. Water Resour. Plan. Manag.* 108, 243–256.

Clayton, C.R.I., Xu, M., Whiter, J.T., Ham, A., Rust, M., 2010. Stresses in cast-iron pipes due to seasonal shrink-swell of clay soils. *Proc. Inst. Civ. Eng. - Water Manag.* 163, 157–162.
<https://doi.org/10.1680/wama.2010.163.3.157>

Colin, X., Audouin, L., Verdu, J., Rozental-Evesque, M., Rabaud, B., Martin, F., Bourguine, F., 2009. Aging of polyethylene pipes transporting drinking water disinfected by chlorine dioxide. I. Chemical aspects. *Polym. Eng. Sci.* 49, 1429–1437. <https://doi.org/10.1002/pen.21258>

Davis, P., Silva, D. De, Marlow, D., Moglia, M., Gould, S., Burn, S., 2008. Failure prediction and optimal scheduling of replacements in asbestos cement water pipes. *J. Water Supply Res. Technol.* 57, 239–252.
<https://doi.org/10.2166/aqua.2008.035>

Dingus, M., Haven, J., Austin, R., AWWA Research Foundation., 2002. Nondestructive, non-invasive assessment of underground pipelines. AWWA Research Foundation and American Water Works Association.

Ductile Iron Pipe Research Association, (U.S), 1984. Handbook of ductile iron pipe.

Edil, T.B., Bahmanyar, G.H., 1983. Stresses in buried conduits caused by freezing front. *Cold Reg. Sci. Technol.* 8, 129–137.
[https://doi.org/10.1016/0165-232X\(83\)90004-6](https://doi.org/10.1016/0165-232X(83)90004-6)

Edwards, D.B., Smith, K.M., Duvall, D.P., Grzetic, J.F., 2009. Analysis and testing of freezing phenomena in plastic piping systems 1–6.

- EL-Bagory, T.M.A.A., Younan, M.Y.A., 2016. Crack Growth Behaviour of Pipes Made From Polyvinyl Chloride Pipe Material 1. J. Press. Vessel Technol. 139, 011404. <https://doi.org/10.1115/1.4033124>
- Ellison, D., Spencer, D., 2016. The True Causes of AC Pipe Failures — According to the Data Dan Ellison, P.E. 1 ; and David Spencer, P.E. 2 1 637–647.
- Farewell, T.S., Hallett, S.H., Hannam, J.A., Jones, R.J.A., 2012a. Soil impacts on national infrastructure in the UK. Infrastructure Transitions Research Consortium.
- Farewell, T.S., Hallett, S.H., Truckell, I.G., 2012b. Soil and Climatic causes of water mains infrastructure bursts. 111pp. NSRI Research report. NSRI, Cranfield University, UK.
- Farewell, T.S., Jude, S., Pritchard, O.G., 2018. How the impacts of burst water mains are influenced by soil sand content. Nat. Hazards Earth Syst. Sci. 18, 2951–2968. <https://doi.org/10.5194/nhess-18-2951-2018>, 2018.
- Farrow, J., Jesson, D., Mulheron, M., Nensi, T., Smith, P., 2017. Achieving zero leakage by 2050: the basic mechanisms of bursts and leakage, UK Water Industry Research Limited. London.
- Folkman, S., 2018. Water Main Break Rates in the USA and Canada: A Comprehensive Study 1–49.
- Fuchs-Hanusch, D., Friedl, F., Scheucher, R., Kogseder, B., Muschalla, D., 2013. Effect of seasonal climatic variance on water main failure frequencies in moderate climate regions. Water Supply 13, 435–446. <https://doi.org/10.2166/ws.2013.033>
- Ghabeche, W., Alimi, L., Chaoui, K., 2015. Degradation of Plastic Pipe Surfaces in Contact with an Aggressive Acidic Environment. Energy Procedia 74, 351–364. <https://doi.org/10.1016/J.EGYPRO.2015.07.625>
- Gong, J., Lambert, M., Zecchin, A., Simpson, A., Arbon, N., Kim, Y.-I., 2016. Field

- study on non-invasive and non-destructive condition assessment for asbestos cement pipelines by time-domain fluid transient analysis. *Struct. Heal. Monit.* 15, 113–124. <https://doi.org/10.1177/1475921715624505>
- Gould, S.J.F., Boulaire, F.A., Burn, S., Zhao, X.L., Kodikara, J.K., 2011. Seasonal factors influencing the failure of buried water reticulation pipes. *Water Sci. Technol.* 63, 2692–2699. <https://doi.org/10.2166/wst.2011.507>
- Gould, S.J.F., Davis, P., Beale, D.J., Marlow, D.R., 2013. Failure analysis of a PVC sewer pipeline by fractography and materials characterization. *Eng. Fail. Anal.* 34, 41–50. <https://doi.org/10.1016/j.engfailanal.2013.07.009>
- Goulter, I.C., Kazemi, A., 1988. Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Can. J. Civ. Eng.* 15, 91–97. <https://doi.org/10.1139/l88-010>
- Guoquan, W., Yiaoting, C., 1991. Test methods for gelation of PVC plastisol. *Polym. Test.* 10, 315–324. [https://doi.org/10.1016/0142-9418\(91\)90025-S](https://doi.org/10.1016/0142-9418(91)90025-S)
- Guyer, J.P., 2013. An introduction to pumping stations for water supply systems. Createspace Independent Pub.
- Habibian, A., 1994. Effect of Temperature Changes on Water-Main Breaks. *J. Transp. Eng.* 120, 312–321. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1994\)120:2\(312\)](https://doi.org/10.1061/(ASCE)0733-947X(1994)120:2(312))
- Holder, S.L., Hedenqvist, M.S., Nilsson, F., 2019. Understanding and modelling the diffusion process of low molecular weight substances in polyethylene pipes. *Water Res.* 157, 301–309. <https://doi.org/10.1016/J.WATRES.2019.03.084>
- Hou, Y., Lei, D., Li, S., Yang, W., Li, C.Q., 2016. Experimental Investigation on Corrosion Effect on Mechanical Properties of Buried Metal Pipes. *Int. J. Corros.* 2016. <https://doi.org/10.1155/2016/5808372>
- Hu, Y., Hubble, D.W., 2007. Factors contributing to the failure of asbestos cement

- water mains. Can. J. Civ. Eng. 34, 608–621. <https://doi.org/10.1139/I06-162>
- Hu, Y., Wang, D., Chowdhury, R., 2013. Guidance Manual for Managing Long-term Performance of Asbestos Cement Pipe.
- Iličić, K., 2009. The analysis of influential factors on the frequency of pipeline failures. Water Supply 9, 689–698. <https://doi.org/10.2166/ws.2009.779>
- Kakoudakis, K., Farmani, R., Butler, D., 2018. Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. J. Hydroinformatics 20, 1191–1200. <https://doi.org/10.2166/hydro.2018.152>
- Karpachevskii, L.O., Goroshevskii, A. V., Zubkova, T.A., 2011. Interaction between soils and gas pipelines. Eurasian Soil Sci. 44, 332–339. <https://doi.org/10.1134/S1064229311030045>
- Kettler, A.J., Goulter, I.C., 1985. An analysis of pipe breakage in urban water distribution networks. Can. J. Civ. Eng. 12, 286–293. <https://doi.org/10.1139/I85-030>
- Kimutai, E., Betrie, G., Brander, R., Sadiq, R., Tesfamariam, S., 2015. Comparison of Statistical Models for Predicting Pipe Failures: Illustrative Example with the City of Calgary Water Main Failure. J. Pipeline Syst. Eng. Pract. 6, 04015005. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000196](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000196)
- Kirby, P.C., 1981. PVC PIPE PERFORMANCE IN WATER MAINS AND SEWERS. AGARD Rep. 161–174.
- Kowalska, B., Klepka, T., Kowalski, D., 2016. Influence of chlorinated water on mechanical properties of polyethylene and polyvinyl chloride pipes. Urban Water III 1, 63–74. <https://doi.org/10.2495/uw160061>
- Laucelli, D., Rajani, B., Kleiner, Y., Giustolisi, O., 2014. Study on relationships between climate-related covariates and pipe bursts using evolutionary-

- based modelling. J. Hydroinformatics 16, 743–757.
<https://doi.org/10.2166/hydro.2013.082>
- Lin, P., Zhang, B., Wang, Yi, Li, Z., Li, B., Wang, Yang, Chen, F., 2015. Data Driven Water Pipe Failure Prediction, in: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. ACM, New York, NY, USA, pp. 193–202.
<https://doi.org/10.1145/2806416.2806509>
- Makar, J.M., Desnoyers, R., Mcdonald, S.E., 2001. Failure modes and mechanisms in grey cast iron pipe. Natl. Res. Counc. Canada.
- Martínez-Codina, Castillo, M., González-Zeas, D., Garrote, L., 2016. Pressure as a predictor of occurrence of pipe breaks in water distribution networks. Urban Water J. 13, 676–686.
<https://doi.org/10.1080/1573062X.2015.1024687>
- Mikdam, A., Colin, X., Minard, G., Billon, N., Maurin, R., 2017. A kinetic model for predicting the oxidative degradation of additive free polyethylene in bleach disinfected water. Polym. Degrad. Stab. 146, 78–94.
<https://doi.org/10.1016/j.polymdegradstab.2017.09.020>
- Mordak, J., Wheeler, J., 1988. Deterioration of Asbestos Cement Water Mains, Water Research Council.
- Newport, R., 1981. Factors influencing the occurrence of bursts in iron water mains. Water Supply Manag. 3, 274–278.
- Ofwat, 2018. Final reporting guidance for PR19 2 Exclusions.
- Pelletier, G., Mailhot, A., Villeneuve, J.P., 2003. Modeling Water Pipe Breaks—Three Case Studies. J. Water Resour. Plan. Manag. 129, 115–123.
[https://doi.org/10.1061/\(ASCE\)0733-9496\(2003\)129:2\(115\)](https://doi.org/10.1061/(ASCE)0733-9496(2003)129:2(115))
- Pietrucha-Urbanik, K., 2015. Failure analysis and assessment on the exemplary water supply network. Eng. Fail. Anal. 57, 137–142.

<https://doi.org/10.1016/j.engfailanal.2015.07.036>

- Pozos-Estrada, O., Sánchez-Huerta, A., Breña-Naranjo, J., Pedrozo-Acuña, A., 2016. Failure Analysis of a Water Supply Pumping Pipeline System. *Water* 8, 395. <https://doi.org/10.3390/w8090395>
- Pritchard, O., Hallett, S.H., Farewell, T.S., 2013. Soil Corrosivity in the UK – Impacts on Critical Infrastructure. *Infrastructure Transitions Res. Consort.* 55.
- Pritchard, O.G., Hallett, S.H., Farewell, T.S., 2015a. Soil geohazard mapping for improved asset management of UK local roads. *Nat. Hazards Earth Syst. Sci.* 15, 2079–2090. <https://doi.org/10.5194/nhess-15-2079-2015>
- Pritchard, O.G., Hallett, S.H., Farewell, T.S., 2015b. Probabilistic soil moisture projections to assess Great Britain's future clay-related subsidence hazard. *Clim. Change* 133, 635–650. <https://doi.org/10.1007/s10584-015-1486-z>
- Pritchard, O.G., Hallett, S.H., Farewell, T.S., 2014. Soil impacts on UK infrastructure: current and future climate. *Proc. Inst. Civ. Eng. - Eng. Sustain.* 167, 170–184. <https://doi.org/10.1680/ensu.13.00035>
- Punurai, W., Davis, P., 2017. Prediction of asbestos cement water pipe aging and pipe prioritization using Monte Carlo simulation. *Eng. J.* 21, 1–13. <https://doi.org/10.4186/ej.2017.21.2.1>
- Rajani, B., Kleiner, Y., 2001. Comprehensive review of structural deterioration of water mains: Physically based models. *Urban Water* 3, 151–164. [https://doi.org/10.1016/S1462-0758\(01\)00032-2](https://doi.org/10.1016/S1462-0758(01)00032-2)
- Rajani, B., Kleiner, Y., Sink, J.-E., 2012. Exploration of the relationship between water main breaks and temperature covariates. *Urban Water J.* 9, 67–84.
- Rajani, B., Tesfamariam, S., 2004. Uncoupled axial, flexural, and circumferential pipe & soil interaction analyses of partially supported jointed water mains.

- Can. Geotech. J. 41, 997–1010. <https://doi.org/10.1139/t04-048>
- Rajani, B., Zhan, C., Kuraoka, S., 1996. Pipe-soil interaction analysis of jointed water mains. Can. Geotech. Journal 33, 393–404.
- Rajeev, P., Kodikara, J., Robert, D., Zeman, P., Rajani, B., 2014. Factors contributing to large diameter water pipe failure as evident from failure inspection, in: LESAM 2013 - IWA Leading-Edge Strategic Asset Management. Sydney, Australia, pp. 9–14.
- Rajkolhe, R., Khan, J.G., 2014. Defects, Causes and Their Remedies in Casting Process: A Review. Int. J. Res. Advent Technol. 2, 375–383. <https://doi.org/10.9790/9622->
- Restrepo-Flórez, J.M., Bassi, A., Thompson, M.R., 2014. Microbial degradation and deterioration of polyethylene - A review. Int. Biodeterior. Biodegrad. 88, 83–90. <https://doi.org/10.1016/j.ibiod.2013.12.014>
- Rezaei, H., Ryan, B., Stoianov, I., 2015. Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. Procedia Eng. 119, 253–262. <https://doi.org/10.1016/j.proeng.2015.08.883>
- Rozière, E., Loukili, A., El Hachem, R., Grondin, F., 2009. Durability of concrete exposed to leaching and external sulphate attacks. Cem. Concr. Res. 39, 1188–1198. <https://doi.org/10.1016/j.cemconres.2009.07.021>
- Ruchti, G.F., 2017. Water Pipeline Condition Assessment. American Society of Civil Engineers (ASCE). <https://doi.org/10.1061/9780784414750>
- Scheidegger, A., Leitão, J.P., Scholten, L., 2015. Statistical failure models for water distribution pipes – A review from a unified perspective. Water Res. 83, 237–247. <https://doi.org/10.1016/j.watres.2015.06.027>
- Scheidegger, A., Scholten, L., Maurer, M., Reichert, P., 2013. Extension of pipe failure models to consider the absence of data from replaced pipes. Water Res. 47, 3696–3705. <https://doi.org/10.1016/j.watres.2013.04.017>

- Silva, D. De, Davis, P., Burn, L.S., Ferguson, P., Massie, D., Cull, J., Eiswirth, M., Heske, C., 2002. Condition Assessment of Cast Iron and Asbestos Cement Pipes by In-Pipe Probes and Selective Sampling for Estimation of Remaining Life, in: Fischer, G. (Ed.), Title of Host Publication 20th International No-Dig Conference and Exhibition. Copenhagen, Denmark, pp. 1–13.
- Thacker, K.B., Scholar, P.G., 2015. Analysis of parameters for casting ductile iron pipe-A Review 3, 496–503.
- Trew, J.E., Tarbet, N.K., De Rosa, P.J., 1995. Pipe materials selection manual (1995 edition). Open Library. WRc plc, Marlow.
- Twort, A.C., Ratnayaka, D.D., Brandt, M.J., 2001. Water Supply, 5th ed. Butterworth-Heinemann and IWA Publishing.
- Utility Provider, 2018. Drinking water network failure data collected between 2005 - 2017.
- Van Erp, J., Huisman, W., Vande Walle, G., 2015. The Routledge handbook of white-collar and corporate crime in Europe. Routledge.
- Volk, C., Dundore, E., Schiermann, J., Lechevallier, M., 2000. Practical evaluation of iron corrosion control in a drinking water distribution system. Water Res. 34, 1967–1974. [https://doi.org/10.1016/S0043-1354\(99\)00342-5](https://doi.org/10.1016/S0043-1354(99)00342-5)
- Vreeburg, J.H.G., Vloerbergh, I.N., Van Thienen, P., de Bont, R., 2013. Shared failure data for strategic asset management. Water Supply 13, 1154–1160. <https://doi.org/10.2166/ws.2013.111>
- Wasim, M., Shoaib, S., Mubarak, N.M., Inamuddin, Asiri, A.M., 2018. Factors influencing corrosion of metal pipes in soils. Environ. Chem. Lett. 16, 861–879. <https://doi.org/10.1007/s10311-018-0731-x>
- Wengström, T., 1993. Comparative analysis of pipe break rates: a literature review. Chalmers University of Technology.

- Wols, B. A., Van Thienen, P., 2014. Modelling the effect of climate change induced soil settling on drinking water distribution pipes. *Comput. Geotech.* 55, 240–247. <https://doi.org/10.1016/j.compgeo.2013.09.003>
- Wols, B.A., Van Daal, K., Van Thienen, P., 2014. Effects of climate change on drinking water distribution network integrity: Predicting pipe failure resulting from differential soil settlement. *Procedia Eng.* 70, 1726–1734. <https://doi.org/10.1016/j.proeng.2014.02.190>
- Wols, B.A., Van Thienen, P., 2016. Impact of climate on pipe failure: Predictions of failures for drinking water distribution systems. *Eur. J. Transp. Infrastruct. Res.* 16, 240–253. <https://doi.org/10.1109/EMBC.2015.7319210>
- Wols, B.A., Van Thienen, P., 2014. Impact of weather conditions on pipe failure: a statistical analysis. *J. Water Supply Res. Technol.* 63, 212–223. <https://doi.org/10.2166/aqua.2013.088>
- Wols, B.A., Vogelaar, A., Moerman, A., Raterman, B., 2019. Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. *Water Supply* 19, 404–416. <https://doi.org/10.2166/ws.2018.085>
- Yahaya, N., Lim, K., Noor, N., Othman, S., Abdullah, A., 2011. Effects Of Clay and Moisture Content on Soil-Corrosion Dynamic. *Malaysian J. Civ. Eng.* 23, 24–32.

3 Investigating the Effects of Operational Factors on Pipe Failure in Clean Water Networks

Objective 2	<div><div>Chapter 3: Investigating the effects of operational factors on pipe failures in clean water networks</div><div><ul style="list-style-type: none">▪ Understand the UK WDN data by investigating how operational variables influence pipe failure. Use understanding of Objective 1 to transform data.▪ Identify the most predictive operational variables to use when developing the pipe failure model.</div></div>
-------------	---

Based on the findings from Chapter 2, the following chapter investigates the effect of operational variables on pipe failure frequency in the UK WDN case study. In this investigation, operational variables identified as important from industry professionals have been used, variables that are not often investigated in literature. This investigation fulfils Objective 2 which extends the current understanding of how operational factors effect pipe failure frequency. This chapter is currently under review in a relevant Journal.

Investigating the Effects of Operational Factors on Pipe Failure in Clean Water Networks

Neal Andrew Barton^a, Stephen Henry Hallett^{a*} and Simon Richard Jude^a

^a School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

**Corresponding author: Email: s.hallett@cranfield.ac.uk; Tel: +441234752978*

Abstract

Statistical pipe failure models benefit from a knowledge of the mechanisms that cause pipe failure. Using this knowledge will help build more reliable and accurate models, which can aid decision-makers. Operational factors can have a marked influence on pipe failure rates. In this study, a range of operational factors are investigated, including spatio-temporal relationships between failures, the effects of joints, the influence of large water users, flushing events, and different water sources. A mixed-method approach has been used, including failure frequency, spatial relationship, and Dijkstra's path. There is a spatial relationship between failures, yet the relationship between the time of nearby failures is more ambiguous. Water source and, in particular, cold water from reservoirs has a marked effect on iron pipes. In this data set, flushing events, pipe joints, and large users have limited effect on pipe failure rates, yet there were some data limitations. This understanding can be used in the development of statistical models, providing more accurate results which can lead to more informed management decisions to reduce pipe failures.

Key Words: Infrastructure planning; water supply; data analytics; machine learning; smart networks.

Highlights:

- Investigates unique operational factors influencing pipe failures.
- Each factor is discussed by five common pipe materials.
- Methods include failure rates, proximity analysis and Dijkstra's algorithm.
- The findings can help to further the development of pipe failure models.

3.1 Introduction

The reliability of Water Distribution Networks (WDN) is essential to ensure access to continuous drinking water for billions of people. Pipe failures interrupt water supply, lead to potentially severe damage to infrastructure, and waste clean water. As an area of concern for asset management, there is a need to reduce pipe failures and improve serviceability. Asset managers and data scientists are establishing quantitative tools, particularly statistical models, to predict pipe failures to help proactively manage the Water Distribution Network (WDN). Statistical models use historical data correlated with recorded pipe failures to predict and understand potential failures in the future (Scheidegger et al., 2015). Infrastructure data can be incomplete, inconsistent, or inaccurate, it can lead to errors in predictive models (Barton et al., 2019). An improved understanding of pipe failures and the factors that cause them can thus facilitate an improved modelling approach and enable more informed decisions. The historical data required will ideally comprise those factors influencing pipe failure, typically categorised as pipe-intrinsic, environmental, or operational. Detailed distinctions of the various factors have been investigated in a complementary study by the authors (Barton et al., 2019).

Recently pipe failure data has become more widely available with the advance of data driven insights. More data correlated with pipe failures is enhancing the predictive capabilities of statistical models, however it is important to understand how each variable influences pipe failures. On the broader literature, studies have focussed on dynamic weather factors, including temperature, frost, soil moisture deficit, rain and wind (Barton et al., 2020; B. . Wols and van Thienen, 2014; Wols et al., 2019) and seasonality (Gould et al., 2011; Młyński et al., 2021), and on pipe-intrinsic factors such as material, age and diameter, since these are typically present within most pipe failure datasets (Alizadeh et al., 2019; Wang et al., 2009). Operational factors are also known to influence pipe failure frequency, and some factors have been explored. Pipe pressure has been extensively discussed in the literature, reporting that high pressure results in higher failure frequencies (Farmani et al., 2017; Motiee and Ghasemnejad, 2019; Snider and McBean, 2018; Wu et al., 2021). Rezaei, Ryan and Stoianov (2015) studied the effects of cyclical pipe pressure on a WDN and discovered fatigue failure in pipes from cyclical loading caused by consumer patterns and network operations (pressure management). In other literature Cathodic protection, or lining or coating of pipes helps to protect pipes against corrosion and failure from a reduced resilience (Farrow et al., 2017). Schuster and McBean (2008) observed that cathodic protection reduced the probability of pipe failure over a 10 years period compared to pipes without. Kleiner and Rajani (2012) reported pipe failure models to be influenced by the number of previous failures, suggesting that more failures could be representative of a pipe with reduced resilience. De Oliveira, Garrett, et al. (2011) observed a spatio-temporal relationship between failures and reported that localised influences such as management practices or environmental conditions may influence the resilience of the network and cause higher failure frequencies in certain areas. Network conditions can also lead to higher failure rates, such as the number of valves in a network that can, where more valves can protect critical pipe segments (Liu and Kang, 2021), or the type of joint, where studies show increase failure rates, particularly in PVC pipes (Arsénio, Pieterse, et al., 2013); further information on different material types is needed.

These studies have provided useful insight into the influence of operational factors. However, there is a need to investigate the wide range of operational factors that may potentially influence pipe failures, to incorporate correlated variables into pipe failure models and improve predictions. With a focus on improving predictive pipe failure modelling, this study contributes to the broader understanding of operational impacts on pipe failures by considering some of the alternative operational factors identified by water industry experts as potentially influential. The operational factors considered in this study offer further understanding of existing knowledge on the spatiotemporal relationship between failures (De Oliveira et al., 2011b; Goulter and Kazemi, 1988) the effects of joints (Arsénio et al., 2013; Dingus et al., 2002; Kirby, 1981), and different water sources (Farrow et al., 2017), but introduces novelty through investigating individual material types; including Iron, Steel and Ductile (SDI), Asbestos Cement (AC), Polyvinyl Chloride (PVC) (collectively Unplasticised, Post Chlorinated and Molecular Orientated Polyvinyl Chloride) and Polyethylene (PE) (medium and high density). Further contribution is made by exploring operational factors not yet seen in literature, including large water consumers and operational flushing events.

3.2 Methodology

3.2.1 Data

The data examined in this study, collected from a large UK WDN, contains failure record details recorded between 2005 and 2018. A summary of the network data is presented in Table 3-1.

Table 3-1 WDN data collected between 2005 and 2018

Feature	Iron	Asbestos Cement (AC)	Steel and Ductile Iron (SDI)	Polyvinyl Chloride (PVC)	Polyethy lene (PE)
Length of Pipe ¹	11,382	7,090	1,824	5,854	12,274
Failure rate (failures/km/year)	0.27	0.17	0.05	0.17	0.03

¹ Length of network operational by end of December 2018.

Failures are recorded as reactive or proactive. Reactive failures are those failures fixed by repair teams after a failure incidence has been observed and reported. Reactive failures are more likely to be recorded nearer to the date when the failure happened. Proactive failures are those that are found by maintenance teams actively walking the network in search for legacy leaks and water loss. These many have occurred many weeks or months before they are found; therefore, the time when they are found obfuscates the reasons for failure. Reactive pipe failures were used in this study. Spatial data processing was undertaken using ArcGIS (version 10.6), while data assimilation, processing, analysis, and visualisation were completed using R software (version 4.0.0). A summary table of the variables used in this study are presented in Table 3-2.

Table 3-2 Variables considered for the analysis, description, and type.

Variable	Description	Type
Material	Pipe materials – Iron, AC, SDI, PVC, PE.	Categorical
Length	Total pipe length	Numeric (km)
Large Users	Location	Numeric (spatial)
Pipe failure	Location and time of pipe failure	Numeric (spatial/time)
Water Source	The originating source of water, divided into three categories: Groundwater (GW), Surface Water (SW) and Mixed Water (MW).	Categorical
Pipe joints	Location	Numeric (spatial)
Flushing events	Location and date	Numeric (spatial)

3.3 Methods

The analysis undertaken utilises a range of different methods, which are explained for each operational factor. The dataset was partitioned initially into the five main pipe material types to address the performance of each material type and their unique response to the various operational factors.

3.3.1 Spatiotemporal Analysis

To evaluate the influence of spatial distance on pipe failures, we calculate the distance between each failure to the nearest failure using Dijkstra's algorithm, as proposed by De Oliveira et al. (2011a). Dijkstra's algorithm calculates the shortest distance between each vertex (failure) along an edge (the pipe network) from the source vertex $d[s] = 0$ to all subsequent vertices $d[u] = \infty$. The new vertex u is taken as the shortest distance d and this is then taken as the new frontier vertex, and repeated. This is represented mathematically as:

$$d[u] + w(u, v) < d[v] \text{ then } d[v] = d[u] + w(u, v) \quad \text{Eq.3-1}$$

where $d[u]$ is the distance between each vertex, and $w()$ represents the weight of the edge u, v (Ortega-Arranz et al., 2014; Sunita and Garg, 2018). This approach ensures that the network constraints were considered. The difference in time between the first failure and the closest failure to the first used to determine if the two failures occurred within a short time frame of one another.

3.3.2 Joint Failures

The number of failures associated with each joint is calculated using proximity analysis. A 1 m buffer was created around each of the 292,655 joints included in the WDN. An acceptable tolerance for Global Positioning System (GPS) error, as well as the total number of failures within each buffer was then summed. Since no failures were located directly on a joint, the material type recorded for each failure was taken from that recorded in the field, and therefore the pipe where the failure was discovered; the authors are aware that many joints will join two different materials and accept this as a limitation to the study due to the inherent limitations in the data. However, this is a common circumstance in WDN datasets.

3.3.3 Large Water Users, Flushing Events and Water source

Proximity to large users, flushing events and water source origin are all calculated as an estimated standardised rate of failure y , quantified as the number of failures f experienced by total cohort pipe length l and the relevant observation period T , given as $y = \frac{f/l}{T}$ and presented as bursts/1000 km/year. By assuming a Poisson distribution (Asnaashari et al., 2013; Wols et al., 2019) of the data the variance $V(X) = \sigma^2$ is estimated. The dependence of the weather variables has been determined using Pearson's rank correlation coefficient with a significance value of 0.05. Large users in this instance are defined as users that consume greater than 10,000 m³ /year or 23 m³/day, as determined by the tariff, maximum daily demand charge.

3.4 Results and discussion

3.4.1 Spatiotemporal Relationship

One of the first notable spatio-temporal failure investigations were undertaken by Clark et al. (1982), who observed that consecutive failures had a spatiotemporal relationship with the first. Goulter and Kazemi (1988) later established that 22 % of all pipe failures occurred within one meter and 46 % within 20 m of a prior failure. Of those within one meter of one another, 42 % occurred within one day of the initial failure. In a later study by Kimutai et al. (2015) the authors noted a similar approach calculating the spatial relationship of pipe failures, and reported that 33 % of breaks that occurred after the incidence of a break were within 10 m of the first break. This relationship suggests that the first failure influences preceding failures. This has been attributed to the following reasons:

- The pipe is already weakened through corrosion or deterioration, for example from proximal environmental conditions, and is likely to be within a similar region of the pipe; hence re-pressurisation after pipe repair can instigate further failure.
- Soil disturbance from repairing the first failure from excavation (exposure of the pipe during winter and cold temperatures $< 3^{\circ}\text{C}$) and heavy machinery compacting the soil.
- Erosion of bedding from the pressurised water leaking from the first failure.
- Change in surrounding soil moisture content resulting in differential soil movement from the undisturbed soil (Farrow et al., 2017).

De Oliveira et al. (2011) suggested the early approaches were inaccurate since there was a high level of imprecision within the break locations, the studies were not network constrained, and the local nature of the clusters appeared as random events on a larger scale. To improve on these limitations, De Oliveira et al. (2011) used a larger dataset and Dijkstra's algorithm, which constrained the linked failures to the network, finding that only 6 % of failures were sited within one meter of a previous failure, but that 56 % of failures were within 20 m of another.

The results of the spatial analysis for the UK WDN are shown in Figure 3-1. For clarity, we present only those failures within 20 m of each other only, since this is the furthest distance between failures to be considered 'related', consistent with existing literature (De Oliveira et al., 2011). A full view of all failures up to 1 km apart is presented in Figure B-1 of Appendix B. The results show that approximately 6 % of all pipe failures occur within 1 m of each other whilst approximately 40 % of failures occur within 20 m. These results are lower than existing findings by De Oliveira et al. (2011) (6 % and 56 % respectively) and Goulter and Kazemi (1988) (22 % and 46 %, respectively). However, as with previous literature, this shows a potential spatial relationship in the distance between pipe failures. The use of Dijkstra's algorithm proved a valuable means to determine the distance between failures in a network and represents a more accurate record of the distance between failures constrained by a network.

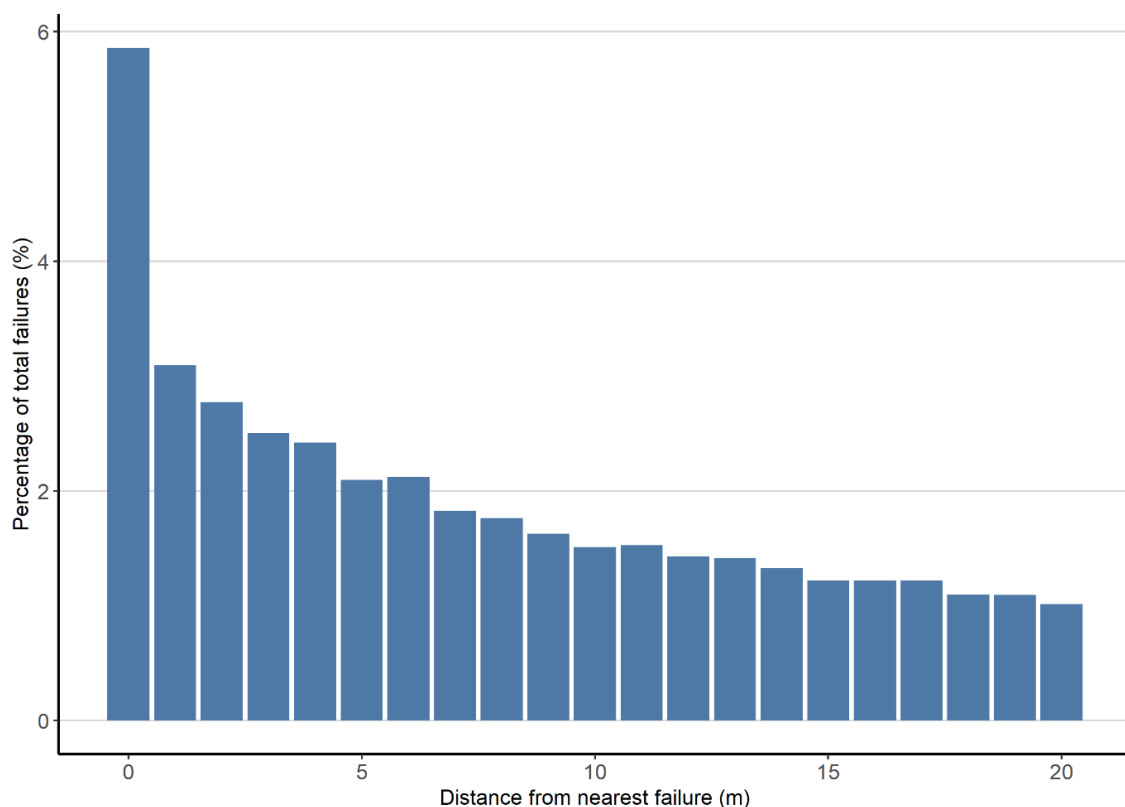


Figure 3-1 Spatial analysis of all pipe failures within 20 m of each other. Data from UK water company for the operational period from 2005 to 2018.

It is considered that a second failure within a close timeframe of the first could potentially arise as a result of soil disturbance or pressure changes occurring from the repair of the first pipe failure (Farrow et al., 2017). Conversely, failures within close proximity to another, but observed a considerable time after the first event are more likely to be associated with the general condition and deterioration of the pipe. To understand which is more prevalent in this WDN, the time between failures is compared, again using failures that are considered to be spatially related (i.e. those within 20 m of each other), which is comparable with previous literature, and a pragmatic approach considering conditions affecting pipe failure can vary greatly across the length of pipes (Andreou et al., 1987a). The time between failures is considered here up to one month apart since anything longer than this was considered unlikely to be related. The results of this analysis are presented in Figure 3-2, revealing that approximately 15 % of all failures within 20 m of another failure, occurred within 30 days. We look further by observing failures occurring within 20 m of another failure, and within one day to compare against the results of Goulter and Kazemi (1988), who observed that 42 % of failures within 20 m occurred within one day. The results in this study show that only approximately 4 % of failures within 20 m occur within one day, which is significantly lower than previously reported. Furthermore, as a percentage of the whole network, failures that occurred within 20 m of one another and within 30 days of one another accounted for only ~ 0.2 %.

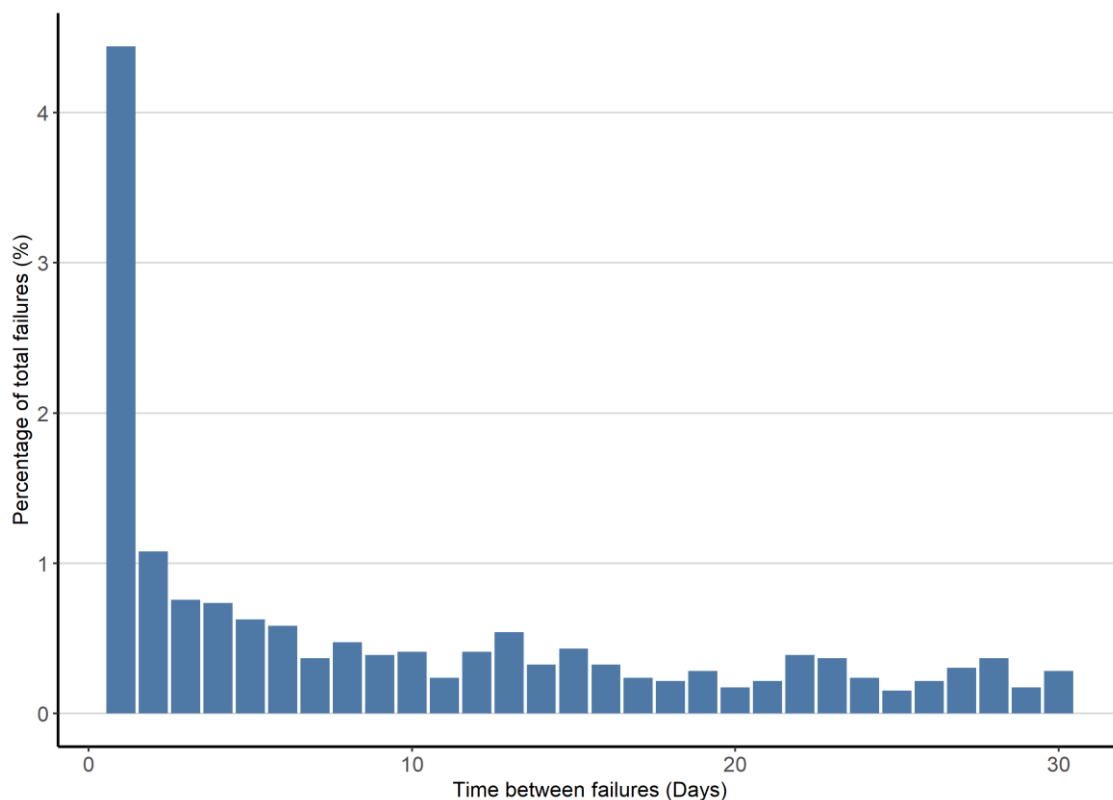


Figure 3-2 Temporal analysis of all pipe failures within 20 m of each other and up to 30 days separation time. Data from UK water company for the operational period from 2005 to 2018.

Distance between failures can thus be considered as a useful variable to consider in pipe failure models, serving as a potential proxy for pipe deterioration; however, the Number Of Previous Failures (NOPF) or recency of failure metrics have also been used successfully in this respect (Kleiner and Rajani, 2012). From this WDN data, the relationship of time between failures is more ambiguous, with only a small percentage of total WDN failures occurring within one day.

3.4.2 Joint Failures

Industry professionals identify that pipe failures frequently occur at joints, something that is also observed in the literature (Arsénio et al., 2013; Dingus et al., 2002; Kirby, 1981). Of particular interest is that PVC pipes have previously shown high joint failure, an average of 16 % of all failures (Arsénio et al., 2013)

because of push-fit joints, 11 % because of poor joint construction and insufficient solvent and 4 % because of surplus solvent (Kirby, 1981). High joint failure is also associated with iron pipes, potentially supporting the notation that ridged joints, typical of iron pipes, are more vulnerable to failing (Barton et al., 2019; Farrow et al., 2017). In response to this, some pipe failure models have used the number of joints within a length of pipe as a covariate (Robles-Velasco et al., 2020). However, determining joint failures is difficult, especially where data on the reason for failure is not recorded.

In the UK WDN dataset, no specific data on type of failure was recorded, and no failures were located directly on a joint (potentially due to GPS error or failing to record an accurate location of the failure). Therefore, we use a simple spatial analysis to understand the association between joints failures, whereby a 1 m buffer is placed around each of the 292,655 joints in the WDN (1 m is considered an acceptable tolerance for GPS error) and the number of failures within each buffer are summed. It is perceivable that joint link two material types, yet in this study we use only the material of the pipe recorded at the time of failure.

Table 3-3 shows the results and suggests the percentage of failures associated with joints in this WDN are far fewer than those reported in previous literature. For example, PVC joint failures only accounted for 2.52 % of all PVC failures, a considerably lower percentage than 14 % and 16 % observed in other studies. The results show the highest failure percentage by material type for PE of 4 %. It was noted that PE has the highest number of failures related to joint failures due to poor electrofusion joint construction, a process which requires specialist training (Barton et al., 2019). Iron, AC and SDI all show low percentages of pipe failures. Overall joint failures accounted for approximately 3 % of all network failures, which is considered low. Consequent to the limitations noted here, it can be deduced that data collection should include the type of failure, and the materials and diameters of the adjoining pipes.

Table 3-3 Percentage of failures within 1 m of a joint for five common pipe materials. Percentage calculated as a percentage of failures for that specific material.

Material	Iron	Asbestos Cement	Steel and Ductile Iron	Polyvinyl Chloride	Polyethylene
Percentage (%)	2.57	1.82	3.70	2.52	4.11

3.4.3 Proximity to Large Water Users

Large users consume more water at any given time than other users in the network, and it is thought that this causes a higher frequency of pipe failures due to the cyclical pipe pressure variations associated with higher flow rates (Farmani et al., 2017). To see if this is true for the UK WDN, the number of large users in each District Metered Area (DMA) were summed, and the failure frequency for each DMA was observed. Figure 3-3 shows the failure rates categorised by large user. The results presented suggest there is no apparent variation in pipe failure frequency for any material when more large users are present within a DMA, compared to when few large users are present within the DMA. This finding is of consequence for future design analysis and pipe failure modelling approaches.

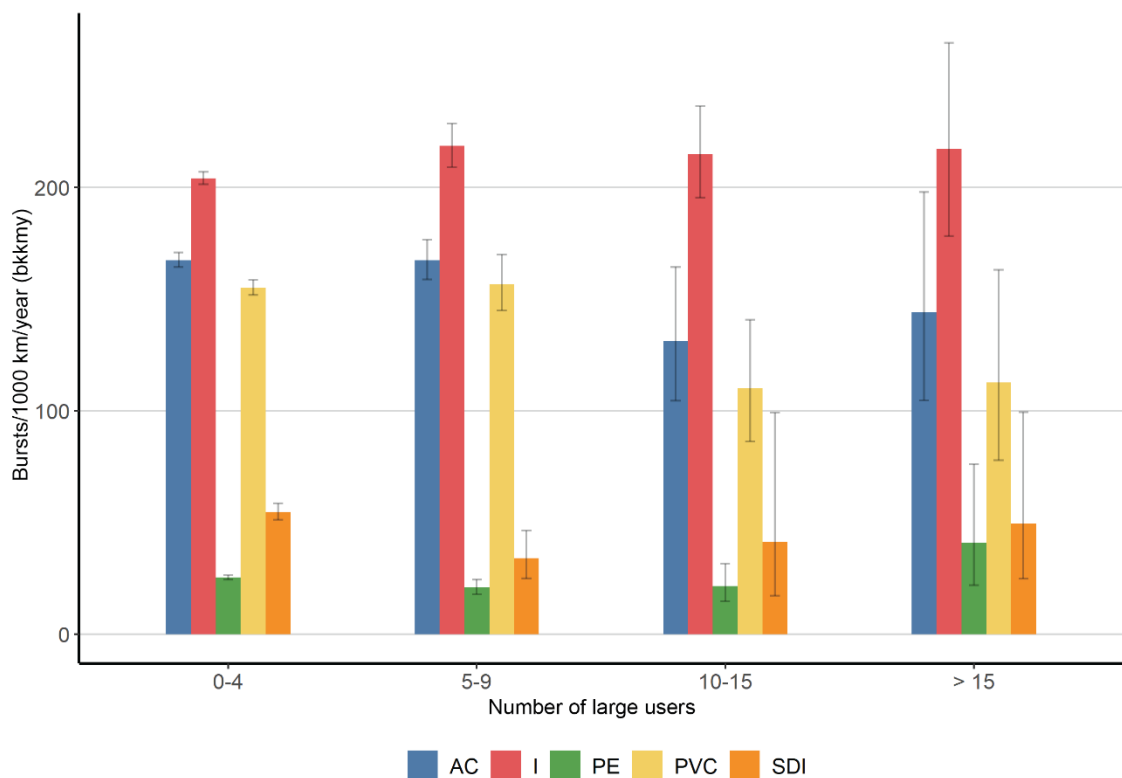


Figure 3-3 Estimated pipe failure rates (bursts/1000 km/year) in DMAs with large users, separated by five common material type. The error bars show a 95 % confidence interval.

3.4.4 Flushing Events

Operational non-repair activities are reported to cause pipe failures (Røstum, 2000). Flushing pipes is a common non-repair activity and involves forcing water intermittently through pipes at a velocity of between 0.70 and 1.40 m/s to clean out sediment build-up. Wu and Liu (2017) observed an increase in flow rates after flushing events, potentially suggesting a failure in the network. To understand if flushing events result in higher failure rates, an analysis was completed on a limited available dataset collected between 2015 and 2019. The data was cleaned due to erroneous data, reducing the data set by approximately 9 %. Of the 19,598 rows of data, 2,781 events were removed due to inaccurate dates, 1,078 rows of data were further removed for not accurately recording the asset number of the pipe, 24 rows were removed for not accurately recording the

material of the pipe, leaving 15,715 instances. This flushing event data set was then merged with the original pipe data (reduced temporally to dates between 2015 and 2019). Since failures are suggested to be delayed from the flushing events, the analysis associated with a pipe failure within two weeks as potentially related.

Figure 3-4 shows the failure frequency associated with the flushing events for each material type. Except for PE, the failure frequency is higher when there is a flushing event; however, the confidence interval shows greater uncertainty in the flushing event data, since there was significantly less pipe used to standardise the results due to the limited data set. These failures may also be related to other causal factors, and so further investigation is required, particularly with environmental factors and seasonality which is outside the scope of this study. This analysis cannot determine if flushing pipes contribute to increased failure frequencies in another ways, for example over several years, or understand the effects of different flushing pressures, since this information is unavailable.

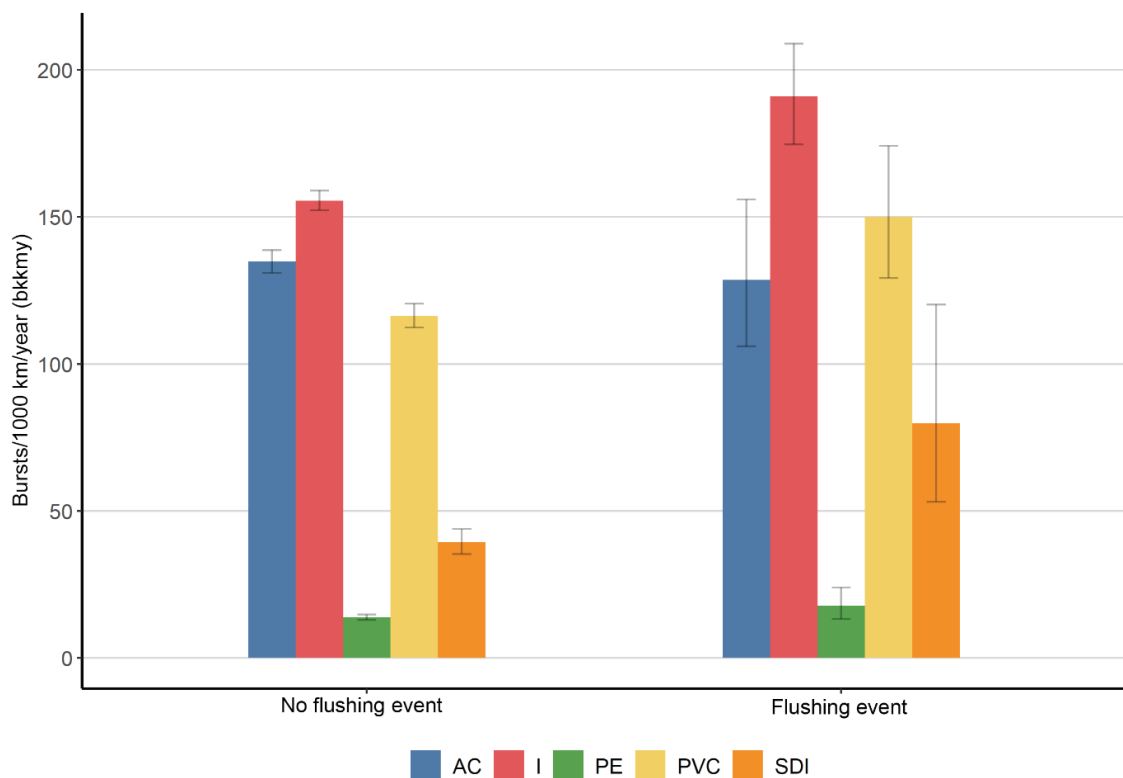


Figure 3-4 Estimated pipe failure rates (bursts/1000 km/year) within two weeks of a flushing event and without flushing events, separated by five common material type. The error bars show a 95 % confidence interval.

3.4.5 Water Source

The effects of water source on pipe failures have been discussed in the literature. Farrow et al. (2017) suggest higher failure frequency can occur in pipes that convey water from reservoirs during the winter, since the water temperature can drop down to close to freezing and in the summer in the high teens, resulting in a large differential temperature. Similarly, mixing water from groundwater and surface waters can introduce temperature shocks to the network. To evaluate the difference in failure frequency between water sources, we analysed the failure rates of different pipe materials according to three primary water sources, groundwater, surface water (reservoirs) and mixed water, to reveal any difference in failure frequency between them. Figure 3-5 identifies that failure rates for iron, steel and ductile iron and asbestos cement are higher when the water is

conveyed from surface water, such as reservoirs, and supports the notion that water from reservoirs results in higher failure rates (Farrow et al., 2017). These materials, especially metal, are efficient thermal conductors, whereas plastic pipes, PVC and PE are not, and therefore show little difference in failure rate.

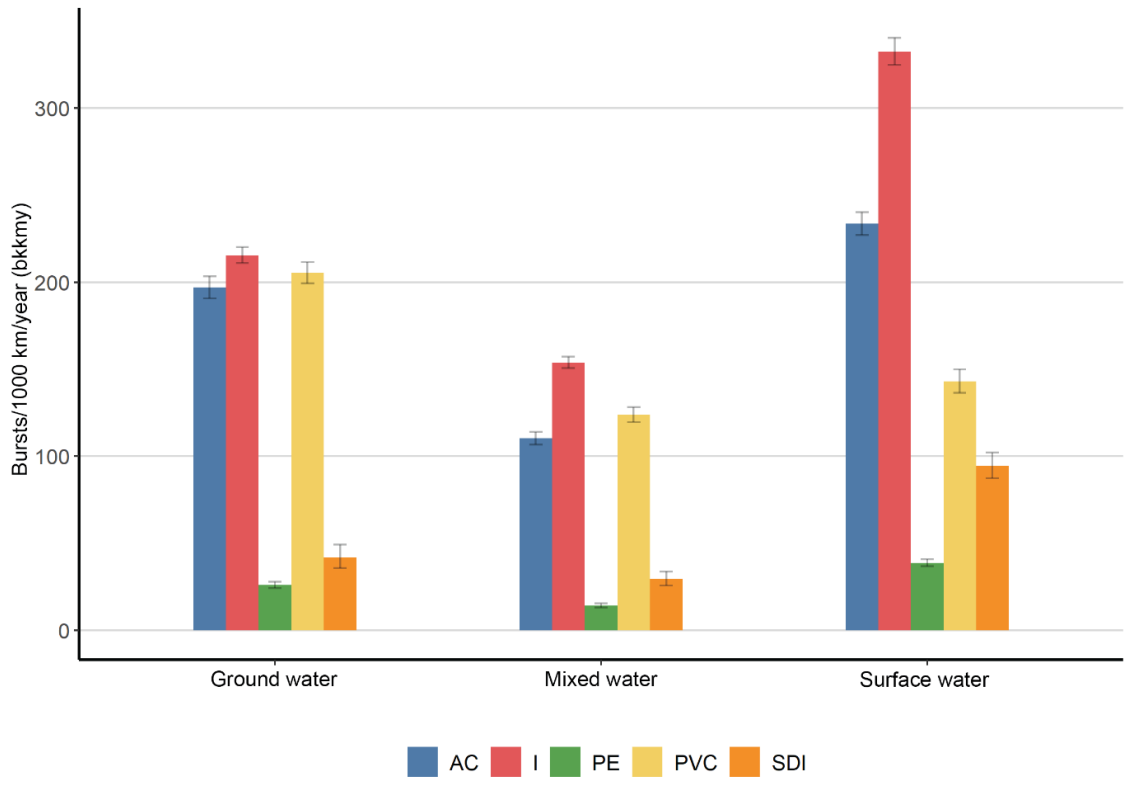


Figure 3-5 Estimated pipe failure rates (bursts/1000 km/year) by water source, separated by five common material type. The error bars show a 95 % confidence interval.

3.5 Conclusion

This study has investigated some of the principal operational factors identified by professionals that could influence the failure rates of different pipe materials in a UK WDN. The methodology used various approaches to identify each factor separately for five commonly used pipe materials. The study is important since it extends existing literature for some factors by further understanding the influence on different material types, and by considering some alternative factors not yet

considered. The two most influential operational factors discussed in this study include the geographical distance between failures and the water source. There is limited evidence in this WDN dataset to suggest that joint failures, proximity to large users and flushing events result in higher failure rates. However, further investigation on different WDN datasets is required to be certain. The results suggest that distance between failures is important and supports findings in the literature, although conversely to other studies, the failures are not as closely linked in time for this UK WDN. The influence of water sources suggests that surface water results in a higher frequency of failures, potentially a result of surface water is colder during the winter.

The results contribute to current understanding of different operational factors and how they influence pipe failures. This information can be used to improve modelling approaches to predict pipe failures, leading to more reliable models. In turn, managers can use improved models to make more informed management decisions and build more reliable networks that can reduce pipe failures and water loss.

3.6 Author Contributions

Neal Barton: Conceptualisation, Methodology, Investigation, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualisation and Project administration. Stephen Hallett: Writing – Review and Editing, Supervision, Funding Acquisition. Simon Jude: Review and Editing, Supervision, Funding Acquisition.

3.7 References

- Alizadeh, Z., Yazdi, J., Mohammadiun, S., Hewage, K., Sadiq, R., 2019. Evaluation of data driven models for pipe burst prediction in urban water distribution systems. *Urban Water J.* 16, 136–145. <https://doi.org/10.1080/1573062X.2019.1637004>

- Andreou, S.A., Marks, D.H., Clark, R.M., 1987. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. *Adv. Water Resour.* 10, 2–10. [https://doi.org/10.1016/0309-1708\(87\)90002-9](https://doi.org/10.1016/0309-1708(87)90002-9)
- Arsénio, A.M., Pieterse, I., Vreeburg, J., de Bont, R., Rietveld, L., 2013. Failure mechanisms and condition assessment of PVC push-fit joints in drinking water networks. *J. Water Supply Res. Technol.* 62, 78–85. <https://doi.org/10.2166/aqua.2013.026>
- Asnaashari, A., McBean, E.A., Gharabaghi, B., Tutt, D., 2013. Forecasting watermain failure using artificial neural network modelling. *Can. Water Resour. J.* 38, 24–33. <https://doi.org/10.1080/07011784.2013.774153>
- Barton, N.A., Farewell, T.S., Hallett, S.H., 2020. Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *npj Clean Water* 3, 31. <https://doi.org/10.1038/s41545-020-0077-3>
- Barton, N.A., Farewell, T.S., Hallett, S.H., Acland, T.F., 2019. Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Res.* 164, 114926. <https://doi.org/10.1016/j.watres.2019.114926>
- Clark, R.M., Stafford, C.L., Goodrich, J.A., 1982. Water distribution systems: a spatial and cost evaluation. *J. Water Resour. Plan. Manag.* 108, 243–256.
- De Oliveira, D.P., Garrett, J.H., Soibelman, L., 2011a. A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage. *Adv. Eng. Informatics* 25, 380–389. <https://doi.org/10.1016/j.aei.2010.09.001>
- De Oliveira, D.P., Neill, D.B., Garrett, J.H., Soibelman, L., 2011b. Detection of patterns in water distribution pipe breakage using spatial scan statistics for

- point events in a physical network. J. Comput. Civ. Eng. 25, 21–30.
[https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000079](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000079)
- Dingus, M., Haven, J., Austin, R., AWWA Research Foundation., 2002. Nondestructive, noninvasive assessment of underground pipelines. AWWA Research Foundation and American Water Works Association.
- Farmani, R., Kakoudakis, K., Behzadian, K., Butler, D., 2017. Pipe Failure Prediction in Water Distribution Systems Considering Static and Dynamic Factors. Procedia Eng. 186, 117–126.
<https://doi.org/10.1016/j.proeng.2017.03.217>
- Farrow, J., Jesson, D., Mulheron, M., Nensi, T., Smith, P., 2017. Achieving zero leakage by 2050: the basic mechanisms of bursts and leakage, UK Water Industry Research Limited. London.
- Gould, S.J.F., Boulaire, F.A., Burn, S., Zhao, X.L., Kodikara, J.K., 2011. Seasonal factors influencing the failure of buried water reticulation pipes. Water Sci. Technol. 63, 2692–2699. <https://doi.org/10.2166/wst.2011.507>
- Goulter, I.C., Kazemi, A., 1988. Spatial and temporal groupings of water main pipe breakage in Winnipeg. Can. J. Civ. Eng. 15, 91–97.
<https://doi.org/10.1139/l88-010>
- Kimutai, E., Betrie, G., Brander, R., Sadiq, R., Tesfamariam, S., 2015. Comparison of Statistical Models for Predicting Pipe Failures: Illustrative Example with the City of Calgary Water Main Failure. J. Pipeline Syst. Eng. Pract. 6, 04015005. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000196](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000196)
- Kirby, P.C., 1981. PVC Pipe Performance in Water Mains and Sewers, in: International Conference on Underground Plastic Pipe ASCE. New Orleans, LA, pp. 161–174.

- Kleiner, Y., Rajani, B., 2012. Comparison of four models to rank failure likelihood of individual pipes. *J. Hydroinformatics* 14, 659–681. <https://doi.org/10.2166/hydro.2011.029>
- Liu, J., Kang, Y., 2021. Segment-based resilience response and intervention evaluation of water distribution systems. *J. Water Supply Res. Technol.* <https://doi.org/10.2166/aqua.2021.133>
- Młyński, D., Bergel, T., Młyńska, A., Kudlik, K., 2021. A study of the water supply system failure in terms of the seasonality: analysis by statistical approaches. *J. Water Supply Res. Technol.* 70, 289–302. <https://doi.org/10.2166/aqua.2021.151>
- Motiee, H., Ghasemnejad, S., 2019. Prediction of pipe failure rate in Tehran water distribution networks by applying regression models. *Water Supply* 19, 695–702. <https://doi.org/10.2166/ws.2018.137>
- Ortega-Arranz, H., Torres, Y., Llanos, D.R., Gonzalez-Escribano, A., 2014. The All-Pair Shortest-Path Problem in Shared-Memory Heterogeneous Systems, in: Jeannot, E., Zilinskas, J. (Eds.), *High-Performance Computing on Complex Environments*. John Wiley & Sons, Ltd, pp. 283–297.
- Rezaei, H., Ryan, B., Stoianov, I., 2015. Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. *Procedia Eng.* 119, 253–262. <https://doi.org/10.1016/j.proeng.2015.08.883>
- 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. *Reliab. Eng. Syst. Saf.* 196, 106754. <https://doi.org/10.1016/j.res.2019.106754>
- Røstum, J., 2000. *Statistical Modelling of Pipe Failures in Water Networks*. Norwegian University of Science and Technology.

- Scheidegger, A., Leitão, J.P., Scholten, L., 2015. Statistical failure models for water distribution pipes – A review from a unified perspective. *Water Res.* 83, 237–247. <https://doi.org/10.1016/j.watres.2015.06.027>
- Schuster, C.J., McBean, E.A., 2008. Impacts of cathodic protection on pipe break probabilities: a Toronto case study. *Can. J. Civ. Eng.* 35, 210–216. <https://doi.org/10.1139/L07-095>
- Snider, B., McBean, E.A., 2018. Improving Time-to-Failure Predictions for Water Distribution Systems using Gradient Boosting Algorithm, in: 1st International WDSA / CCWI 2018 Joint Conference. Kingston, Ontario.
- Sunita, Garg, D., 2018. Dynamizing Dijkstra: A solution to dynamic shortest path problem through retroactive priority queue. *J. King Saud Univ. - Comput. Inf. Sci.* <https://doi.org/10.1016/j.jksuci.2018.03.003>
- Wang, Y., Zayed, T., Moselhi, O., 2009. Prediction Models for Annual Break Rates of Water Mains. *J. Perform. Constr. Facil.* 23, 47–54. [https://doi.org/10.1061/\(ASCE\)0887-3828\(2009\)23:1\(47\)](https://doi.org/10.1061/(ASCE)0887-3828(2009)23:1(47))
- Wols, B., van Thienen, P., 2014. Impact of weather conditions on pipe failure: a statistical analysis. *J. Water Supply Res. Technol.* 63, 212–223. <https://doi.org/10.2166/aqua.2013.088>
- Wols, B.A., Vogelaar, A., Moerman, A., Raterman, B., 2019. Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. *Water Supply* 19, 404–416. <https://doi.org/10.2166/ws.2018.085>
- Wu, Y., Liu, S., 2017. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water J.* 14, 972–983. <https://doi.org/10.1080/1573062X.2017.1279191>
- Wu, Z.Y., Chew, A., Meng, X., Cai, J., Pok, J., Kalfarisi, R., Lai, K.C., Hew, S.F., Wong, J.J., 2021. Data-driven and model-based framework for smart

water grid anomaly detection and localization. J. Water Supply Res. Technol. 00, 1–11. <https://doi.org/10.2166/aqua.2021.091>

4 Using Generalized Additive Models to Investigate the Environmental Effects on Pipe Failure in Clean Water Networks

Objective 2	<div data-bbox="296 533 603 739">Chapter 4: Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks</div> <div data-bbox="603 533 1361 739"><ul style="list-style-type: none">▪ Understand the UK WDN data by investigating how environmental variables influence pipe failure. Use understanding of Objective 1 to transform data.▪ Identify the most predictive environmental variables to use when developing the pipe failure model.</div>
-------------	---

Based on the findings from Chapter 2, the following chapter investigates the effect of environmental variables on pipe failure frequency in the UK WDN case study. In this investigation, typical variables known to affect pipe failure frequency have been chosen, in addition to environmental variables transformed into suitable metrics to extend the existing understanding in literature. This is in partial fulfilment of Objective 2. The chapter is taken directly, without further edits, from the following article published in the journal npj Clean Water, in line with the Creative Commons Attribution 4.0 international license agreement, whilst the supplementary material can be found in Appendix C:

Barton, N.A., Farewell, T.S., Hallett, S.H., 2020. Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. npj Clean Water 3, 31. <https://doi.org/10.1038/s41545-020-0077-3>

Using Generalized Additive Models to Investigate the Environmental Effects on Pipe Failure in Clean Water Networks

Neal Andrew Barton^a, Timothy Stephen Farewell^{a*} and Stephen Henry Hallett^a

^a School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

**Corresponding author: Email: t.s.farewell@cranfield.ac.uk; Tel: +441234752978*

Abstract

Predicting pipe failures using statistical modelling benefits from detailed knowledge of the conditions and circumstances which influence such failures. Incorporating this knowledge into model building improves failure predictions. In this study, we model weather, soil, and hydrogeological variables in a Generalized Additive Model for five common pipe materials separately, using partial dependence plots to understand the partial effects of each variable on pipe failure. We show how severe temperatures are associated with high pipe failure. Cold temperatures and air frost and their interaction with soils represent the key factors for pipe failures during the winter for metal pipes. Warm temperatures, high Soil Moisture Deficit, and soil movement results in higher pipe failures in Asbestos Cement pipes during the summer. Warm temperatures, ground movement and soil wash out, and water demand are key factors for Polyvinyl Chloride pipe failure during the summer. Frost is a key factor influencing Polyethylene pipes during winter. An understanding of the physical principals concerning pipe failures can enable the development of more accurate models,

guiding network management plans to help reduce asset leakage through appropriate interventions.

Key Words: Pipe failure, Generalized additive model, Environment, Soil, Weather.

Highlights:

- Several weather and soil variables are investigated.
- Partial dependence plots are used to visualise partial dependence.
- Monthly failures are predicted for groups of pipes to capture seasonality.
- An evaluation of the results, exploring how weather and soil effect pipe failures.

4.1 Introduction

In the UK, three billion litres of drinking water is lost daily through pipe failures (Environment Agency, 2018). Pipe failures can amass considerable financial costs from wasted water processing, proximal property and infrastructure damage, service interruption for consumers and network repairs. A comprehensive overview of the impacts of water mains failure has been previously undertaken (Farewell et al., 2018). The implications of pipe failure are significant, and with an estimated 35 % future increase in UK water demand by 2050, in addition to water stress from environmental risks under future climates (Pritchard et al., 2014), water companies are being challenged by the industry regulator to reduce water lost through pipe failures.

Clean water assets must be managed appropriately to reduce the impacts of pipe failure. One way to manage pipe failure is through reactive maintenance, performing repairs only once they have been identified (typically when water reaches the surface). Reactive maintenance is preferably performed immediately after or within a very short time of the failure, depending on its nature (loss of water by volume is sometimes used to determine the severity of the pipe failure

and how quickly the repair must be completed). Sometimes, when pipe failures are small they can remain undetected for months or even years, meaning a considerable volume of water can be lost over time (Scheidegger et al., 2015). To develop a cost-effective means of managing and maintaining networks, statistical models can be developed which are derived from the relationship between pipe failures and causal factors (Pelletier et al., 2003). Statistical models can be used to identify pipes potentially at risk of failure, supporting operational management decisions and classifying pipes for replacement. In the last two decades, UK water companies have increased the quantity and quality of network data, enabling a wider range of data correlated with pipe failure to be considered when building statistical models. An understanding of how each variable influence pipe failure is important and recent studies have provided useful insights, exploring environmental effects on pipes (Fuchs-Hanusch et al., 2013; Gould et al., 2011; Kakoudakis et al., 2017; Wols et al., 2019; Wols and Van Thienen, 2014).

Pipe failure mechanisms are unique for different pipe materials. The literature has reported a correlation between seasonal variation and pipe failures, as a consequence of changing weather (temperature, frost and rainfall deficit (RD)) and soil conditions (pH, ground hazards such as shrink swell, texture, moisture content) (Barton et al., 2019; Bruaset and Sægrov, 2018; Gould et al., 2011; Wols et al., 2019; Wols and Van Thienen, 2014). Failures during the winter are typically found in iron and to a lesser extent steel and ductile iron (SDI) pipes, and are associated with cold temperatures (typically below 3 °C), internal water temperature, rapid temperature transit and prolonged periods of frost (Fuchs-Hanusch et al., 2013; Rajani et al., 2012; Wols et al., 2019). During the summer a higher number of pipe failures are typically found in asbestos cement (AC) and polyvinyl chloride (PVC) pipes (Gould et al., 2011; Wols and Van Thienen, 2014) and are associated with temperature and high RD which results in ground movement from soil shrink swell, associated with clay soils. Although not the same, soil moisture deficit (SMD) can be used as a corollary for RD since both

are measures used to understand soil moisture content and the subsequent effects on volumetric expansion and contraction in clay soils (Farewell et al., 2012b). Polyethylene (PE) is affected by environmental factors, but to a lesser extent than with other materials (Barton et al., 2019). These studies provide useful insights on how environmental factors can influence pipe failures; however, many studies focus on only a few variables to explain pipe failures. Due to their interrelated nature, it is important to look at a wide range of soil and weather variables together to explain environmental effects on pipe failures. Furthermore, understanding the effects of the variables simultaneously, considering the regional environmental and network conditions, is important (Heinze et al., 2018; Ratner, 2010). This understanding can be incorporated into a statistical model to improve pipe failure predictions and inform decision-making (Marra and Wood, 2011).

This study seeks to contribute to the wider understanding of the environmental impacts on pipe failures in common pipe materials, including iron, SDI, AC, PVC (collectively unplasticised, post chlorinated and molecular orientated PVC) and PE (medium and high density). This study performs multivariate analysis using generalized additive models (GAM) to understand how the covariates fit the model and affect pipe failure. GAMs are an approach used extensively in environmental modelling and provide great scope to model complex relationships between covariates. We look for variable dependence and interpret the effects of the covariates using partial dependence plots (PDPs). Pragmatically, visual aids can help to determine which variables have the strongest effect, are useful for interactive model building, and are easy to interpret (Fasiolo et al., 2019). We use data collected over 14 years for an entire network (approximately 40,000 km of pipes) provided by a large UK regional water network.

4.2 Methods

4.2.1 Data Assimilation

Spatial data processing was undertaken using ArcGIS (version 10.6) (ESRI, 2019) while data assimilation, processing, modelling and visualisation were completed using R software (version 3.2.3) (R Core Team, 2018). Using ArcGIS, reactive pipe failure records collected onsite were relocated (snapped) to the nearest matching pipes using a distance of 3 m, based on the nearest pipe with equivalent diameter size and material type to those recorded onsite. Where pipe failures did not match a pipe section located within this 3 m distance, the process was repeated again sequentially up to a distance of 1 km until a match was found. The failure data were then joined to the pipe data using geographical reference. Pipes were attributed through spatial intersection with shrink swell, pH and HOST. The MORECS (Hough and Jones, 1997) and summary weather data metrics were derived to comprise variants of the original data. Pipe diameter and age, which were classified into appropriate bands (Table 4-1 and

Table 4-2).

Table 4-1 Classification bands for pipe age.

Age Band	Range
1	< 1880
2	1880 to < 1900
3	1900 to < 1920
4	1920 to < 1940
5	1940 to < 1960
6	1960 to < 1980
7	1980 to < 2000
8	2000 to < 2021

Table 4-2 Classification bands for pipe diameter

Diameter band	Range
1	< 166 mm
2	166 to < 321 mm
3	321 to < 626 mm
4	≥ 626 mm

These data were split into cohorts of pipes based on the following characteristics: material, age, diameter, 1 km grid, soil type (sand, clay, peat, silt and soft). Dividing, or sorting the pipes into homogeneous groups was found useful for analysis as it reduces the computation time required for big data and reduces data uncertainty; a common approach adopted for statistical modelling of pipes (Kleiner and Rajani, 1999; Zhou, 2018). These data were then summarised by week to join with the weekly weather data. The environmental variables considered for analysis are outlined in Table 4-3 along with the thresholds and units considered. The environmental variables have been selected based on a review of relevant literature undertaken (Barton et al., 2019) that identified the major environmental factors and their influence on pipe failure for different materials.

Table 4-3 Variables considered for statistical analysis, description, and units.

Variable	Description	Units
Pipe Failures	Number of Reactive Bursts.	No.
Length of Pipe	Total length of each pipe asset split by asset number.	m
Pipe age	Pipe age categorised into eight categories (see Table 4-1)	Years
Pipe Diameter	Pipe diameter categorised into four categories (see Table 4-2)	mm
Temperature	Temperature change in 1 °C increments. Temperature change in 1 °C over 1 week. (Weekly mean minus the next weekly mean). Temperature change in 1 °C over 2 weeks. (Fortnightly mean minus the next fortnightly mean).	°C
Days Air Frost	Total days air frost in a month.	Days
SMD	SMD levels. SMD change over 1 week. SMD change over 2 weeks.	mm
Soil Texture	Subsoil texture based the dominant class in the soil association at a depth of 70cm. Soils broadly categorised into clay, loam, peat, rock, and sand.	Categories
Shrink swell	Six categorical levels of soil shrink swell potential with a % volumetric expansion < 3, 3-5, 5-12, 12-15 and 12-15* (when soils are drained to 2 m).	Categories1
pH	Dominant average pH value in the soil based on the soil classification.	pH value

Variable	Description	Units
Pipe Failures	Number of Reactive Bursts.	No.
Length of Pipe	Total length of each pipe asset split by asset number.	m
Pipe age	Pipe age categorised into eight categories (see Table 4-1)	Years
Pipe Diameter	Pipe diameter categorised into four categories (see Table 4-2)	mm
Hydrology of Soil Type	HOST classes are based on substrate hydrogeology flow mechanism and is split into 8 sub classes.	Categories ²

¹See Table C-1 Appendix C

²See Table C-2 Appendix C

4.2.2 Statistical Analysis

Data exploration was undertaken to ascertain the effects of environmental conditions of pipe failure used by modelling the variables against the number of pipe bursts (total number of bursts divided by the pipe length) using a GAM and viewing the results of the PDPs. GAMs are similar to Generalized Linear Models (GLM) but differ by relaxing the linear assumption, potentially revealing non-linear relationships and important structure in these data that would otherwise be missed (Wiley and Wiley, 2019). To this extent, a GAM can show either linear, monotonic, or more complex relationships, depending on the way each variable responds to changes in the dependent variables. The GAM performs in this way by extending a GLM to include a smoothing basis function that can measure arbitrarily nonparametric relationships. Categorical data are treated as a linear term without smoothing (Buja et al., 1989). This semi-parametric approach makes GAMs very flexible, easily accommodating different types of data (Wiley and

Wiley, 2019). It should be noted that PDPs are limited in the sense that they only apply in the context of the model used. PDPs may also be misleading in the presence of variable dependence (known as concurvity in GAMs). The computational methods were implemented from the cran repository 'mgcv' package. For the GAM modelling we assume a Poisson distribution since pipe failures are count data. The notation for the GAM smoothing in a Poisson model is as follows (Hastie and Tibshirani, 1986; Wood and Augustin, 2002):

$$g(E(y_i)) = \beta_0 + \beta_j x_{ji} + \dots + f_k(x_{ki}) \dots \quad \text{Eq.4-1}$$

Where $(E(y_i))$ is one of n observations of the response variable, g the Poisson distributed exponential family with the log link function, β_0 is the mean number of observed pipe failures, β_j is the linear term of some predictor covariate x_i and f_k the smoothing term of some non-parametric predictor covariate x_k . to overcome some issues of working with big data (Wood, 2006). From the 'mgcv' package we use 'bam', a less memory intensive version of 'gam' (Wood et al., 2015). A log offset for pipe length was included as an explanatory variable to correct the difference in pipe length. For the smoothing basis function, we use the penalised cubic regression spline. This is beneficial since it lowers computation cost and avoids overfitting by applying a smoothing penalty when estimating the coefficient, generalising the smoothers by shrinking them towards zero. This penalisation approach also helps to generalise the model which can reduce the effects of concurvity and when using 'select = TRUE' can remove variables from the model if they add no value (Wood and Augustin, 2002). We leave the number of knots as default unless residuals are found to be significant, in which case we adjust upwards accordingly. The smoothing parameter estimation was restricted maximum likelihood ('REML'), typically used for smooth components viewed as random effects (Wiley and Wiley, 2019). The notation for the penalised cubic regression is represented for two variables as follows (Wood and Augustin, 2002):

$$f_1(x_1) = \sum_{j=1}^{q_1} \beta_j b_{1j}(x_1) \quad f_2(x_2) = \sum_{j=1}^{q_2} \beta_{j+q_1} b_{2j}(x_2) \quad \text{Eq.4-2}$$

Where q_j represents the penalty matrix, β_j the unknown coefficient to be estimated and $b_j(x_j)$ and the cubic regression spline basis function for f_j of a non-parametric predictor covariate x_j (Wood and Augustin, 2002).

The occurrence of variable dependence (known as concurvity in GAM models) can result in poor parameter estimation and increased confidence intervals, leading to an increased risk of a false statistically significant effect. The use of penalized cubic regression is an approach that helps reduce the effects of concurvity (Buja et al., 1989), nonetheless it is important to check this data. Firstly, we check to ensure that the model has converged, therefore no major errors have occurred. We then use the ‘concurvity()’ function from the ‘mgcv’ package which measures how a smoothed variable can be approximated by another. All categorical variables are grouped together and measured against each smoothed variable resulting in a single value for concurvity. The concurvity output metric presents three scenarios, worst, observed and estimated. The estimate is the most reliable measure for concurvity which returns a value between zero and one, where 0 = no concurvity and 1 = no identifiability between the variables (Wood, 2006). There is no universal criteria for concurvity, but one study on the effects of concurvity suggests values > 0.5 starts to introduce noticeable errors (He, 2004); we adopt this value as the cut off.

The final results are presented as PDPs using the cran ‘mgcViz’ package (Fasiolo et al., 2019). PDPs are a useful model agnostic tool used to perform a graphical exploratory evaluation of the variable effects by showing the change in the mean predicted number of pipe failures as the variable interval changes over its distribution (mean centred since the smoothed variables must sum to zero in a GAM) (Goldstein et al., 2015), accompanied by the 95 % confidence intervals.

The estimated p -value (< 0.01) will be used to determine the significance of the variable effect on pipe failures (see Appendix C.3).

4.2.3 Data Collection

Historical data collected from 2005 to 2018 have been obtained from three sources: network distribution data from a water company, weather data from the UK Met Office and the national soil map-related Natural Perils Directory and LandIS soils data and maps from Cranfield University (Hallett et al., 2017). The network distribution data are for an entire drinking water network which covers a geographical area of 27,476 km² in the UK, comprising of approximately 40,000 km of pipeline. Datasets representing the network contain information regarding each asset including the location of the drinking water distribution network, and the historic burst locations. The different infrastructure variables included in this dataset include length of pipe, diameter size, year of installation, and age of pipe. Two types of bursts are recorded by the infrastructure operator, namely reactive and proactive. Reactive bursts are those responded to once they are reported (by the infrastructure operator or the customer), while proactive bursts are those actively investigated by the operations team or which can be found during a campaign or 'sweep' of an area; thus, the date of repair is temporally unrelated to the date of failure. Only reactive bursts have been included in this analysis, since proactive bursts can manifest for long periods of time before they are finally found and repaired, which makes it hard to ascertain the original date of failure or to link the failure to dynamic temporal environmental variables, so potentially distorting the final results. A summary of the drinking water network used in our analysis is provided in Table 4-4.

Table 4-4 Summary of drinking water network used in the analysis

Material	Installation year	Total network length (km) ¹	Failure rate/km/year
Iron	1881 to 1921	11,382	3.05

Material	Installation year	Total network length (km) ¹	Failure rate/km/year
Asbestos Cement (AC)	1920 to 1941	7,090	2.39
Steel and Ductile Iron (SDI)	1960 to present	1,824	0.75
Polyvinyl Chloride (PVC)	1960 to 2001	5,854	2.40
Polyethylene (PE)	1981 to present	12,274	0.36
Total	1881 to present	38,424	1.86

¹Length of network currently operational by end of December 2018.

Cranfield University's Natural Perils Directory and LandIS datasets have been used to determine the influence of soil on water infrastructure (Hallett et al., 2017). These data include ground movement (shrink swell) and pH presented as 1:250,000 maps based on field data collected between 1939 and 1987. LandIS provides the soil data including soil substrate class and HOST (Boorman et al., 1995) which classifies soil according to its hydrological behaviour. The authors have broadly categorised the 29 UK wide HOST classes into 10 substrate hydrology classes, of which only 8 occurred within the study area (see Table C-2).

The Met Office historical weather data have been collected from the Met Office Rainfall and Evaporation Calculation System (MORECS) and the summary datasets website (Met Office, 2019). MORECS data provide weekly estimates of SMD, hydrologically effective runoff (rainfall) and temperature, expressed on a gridded basis with a spatial resolution of 40 × 40 km.

4.3 Results

The results for the covariate dependence test are shown in Table 4-5. A weak variable dependency is noted between temperature and days air frost; however, no strong variable dependence (> 0.5) was observed. All models converged successfully suggesting the cubic penalised regression approach has generalized the model enough to reduce the effects of concurvity if present. PDPs

show how the mean number of observed pipe failures (y-axis) changes over the variable interval distribution (x-axis). The mean centred number of pipe failures is represented by the red line, whilst the blue dotted lines indicate the 95 % confidence interval. The ticks on the bottom x-axis show the number of observations by the variable interval. The letter on the y-axis represents either a smoothed variable (s) or categorical variable (c), whilst the number reported in parenthesis (e.g., Temp, 8.18) represents the effective degrees of freedom of the smoothed curves and degrees of freedom for the categorical data. The variation in mean predicted pipe failures reveals how influential the variable is on pipe failures. Therefore, we observe variables with the highest increase or decrease in mean prediction as being those which can explain pipe failures. The effects of each variable are discussed further by each material type.

Table 4-5 Concurvity results for Generalized Additive Model Covariates

	Categorical variables	Temperature	Temperature change over one week	Temperature change over two weeks	Days Air Frost	SMD	SMD change over one week	SMD change over two weeks	pH
Categorical variables	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Temperature		1.00	0.02	0.04	0.32	0.13	0.06	0.06	0.00
Temperature change over one week			1.00	0.10	0.02	0.01	0.02	0.01	0.00
Temperature change over two weeks				1.00	0.02	0.01	0.02	0.02	0.00
Days Air Frost					1.00	0.11	0.05	0.05	0.00
SMD						1.00	0.09	0.11	0.00
SMD change over one week							1.00	0.27	0.00
SMD change over two weeks								1.00	0.00
pH									1.00

Partial Dependence Plots (PDPs) show how the mean number of observed pipe failures (y-axis) changes over the variable interval distribution (x-axis). The mean centred number of pipe failures is represented by the red line, whilst the blue dotted lines indicate the 95 % confidence interval. The ticks on the bottom x-axis show the number of observations by the variable interval. The letter on the y-axis represents either a smoothed variable (s) or categorical variable (c), whilst the number reported in parenthesis (e.g., Temp, 8.18) represents the effective degrees of freedom of the smoothed curves and degrees of freedom for the categorical data. The variation in mean predicted pipe failures reveals how

influential the variable is on pipe failures. Therefore, we observe variables with the highest increase or decrease in mean prediction as being those which can explain pipe failures. The effects of each variable are discussed further by each material type.

Figure 4-1 Generalized additive model - partial dependence plots showing the partial dependence of selected variables on Iron pipe failures.

The GAM PDPs for iron pipes are presented in Figure 4-1 and show that all variables included in the model are significant. The highest number of pipe failures can be seen in pipes with a diameter < 166 mm, pipe failures then decrease as the pipe diameter increases. The pipes age does not appear to show a distinct relationship with pipe failure and has high uncertainty in these data suggesting either a small or no effect. Variables leading to the highest change in iron pipe failure includes temperature, days air frost, SMD and diameter band. These variables show a distinct increase in pipe failures over all or a majority of the variable interval. Iron pipe failures increase as the temperature decreases below 10 °C and with more days of air frost in a month up to approximately 15 days, thereafter, reducing as a consequence of fewer observations and a wider confidence interval. Iron pipe failures also increase as SMD increases and is consistent with temperature where pipe failures increase above 15 °C. Soil shrink swell and subsoil type shows a small effect on iron pipe failures but suggested peat and clay soils and soil with a shrink swell potential of 12–15 % volumetric expansion are associated with more pipe failures. Soil pH shows no obvious effect on iron pipe failures which is uncharacteristic since highly acidic or alkaline soils are typically associated with corrosion in metal pipes. However, pH is only one aspect of soil conditions that cause corrosion.

Figure 4-2 Generalized additive model - partial dependence plots showing the partial dependence of selected variables on Asbestos Cement pipe failures.

Figure 4-2 shows the GAM PDPs for AC pipes and indicates higher failure rates for pipes with a diameter < 166 mm and pipes installed between 1940 and 1961. Temperature, SMD and days air frost appear to have the largest effect on AC pipe failures, where more failures occur when temperatures are high (> 15 °C), SMD is high (> 100 mm when soils are drying out) and between 10–15 days air frost in a month; confidence intervals after 15 days increase as the number of observations decrease. Temperature change over one week is not significant while temperature change over two weeks and SMD change over one and two weeks all show a small variation; however, the confidence intervals are large suggesting minimal effect on AC pipe failures. For the soil variables a higher number of AC pipe failures are observed in high soil pH (> 11), in alluvial peats and clays with a large shrink swell potential of 12–15 % volumetric expansion when drained to a depth of 2m. Hydrologically impermeable (soft) soils (hydrology of soil type (HOST) category 8) comprising cover loam, clay with flints or plateau drift and loamy drift shows a higher number of AC pipe failures.

Figure 4-3 Generalized additive model - partial dependence plots showing the partial dependence of selected variables on Steel and Ductile Iron pipe failures.

The results of the GAM PDPs for SDI pipes are presented in Figure 4-3. Temperature, SMD, SMD change over one and two weeks are not significant in the model. Observing the PDPs, the variables revealing the largest effect on SDI pipe failures include days air frost and pipe diameter. As the number of days air frost in a month increases the SDI pipe failures increase, and there is an apparent increase in SDI pipe failures for smaller diameter pipes (small pipes < 166 mm). Pipe age indicated a fall in pipe failures for newer pipe installations; however, the confidence interval is wide. Observing shrink swell, the highest number of failures is noted in soils with low and medium shrink swell potential suggesting that soil movement has limited effect on SDI pipe failures. Peat and clay subsoils show the highest number of SDI failures, although the confidence interval for peat is high suggesting too few observations for this soil category. There are a higher number of SDI pipe failures in soils which are unconsolidated and microporous (HOST category 5), these soils comprise mudstones, soft massive clays, very soft clays and loams. The remaining variables showed a limited effect on the number of pipe failures.

Figure 4-4 Generalized additive model - partial dependence plots showing the partial effects of selected variables on Polyvinyl Chloride pipe failures.

The GAM PDPs for PVC pipes are presented in Figure 4-4. The difference between the diameter bands shows a large uncertainty in the largest pipes > 625 mm due to a small number of observations and shows no obvious variation between other diameter bands to suggest there is any real effect on PVC pipe failures. Pipe age shows little effect on pipe failures evidenced by the small variation between categories. SMD change over two weeks and temperature change over one week shows no significance in the model and looking at the PDPs there is also a limited effect from temperature change over two weeks and SMD change over one week, both revealing little variation in PVC pipe failures and high confidence intervals. Observing soil pH, there is a minor increase in failures for more acidic soils; however, the confidence interval increases, and the number of observations is limited, therefore it is considered that highly acidic or alkaline soils has little effect on PVC pipe. Weather variables with the most effect on PVC failures include temperature and days air frost. Considering temperature, the number of PVC pipe failures increases as the temperature increases, and the same can be seen for days air frost in a month where the number of failures increases as the number of days frost in a month increases. Considering subsoil type, peat has the highest effect on PVC pipe failures, although the confidence interval is large due to a small number of observations. Loam and sand dominant soils show more certainty and have a minor effect on higher pipe failures. Soils with weakly consolidated macroporous soils (HOST category 2) comprising chalk and chalk rubble are also associated with the highest number of PVC failures.

4.3.5 Polyethylene

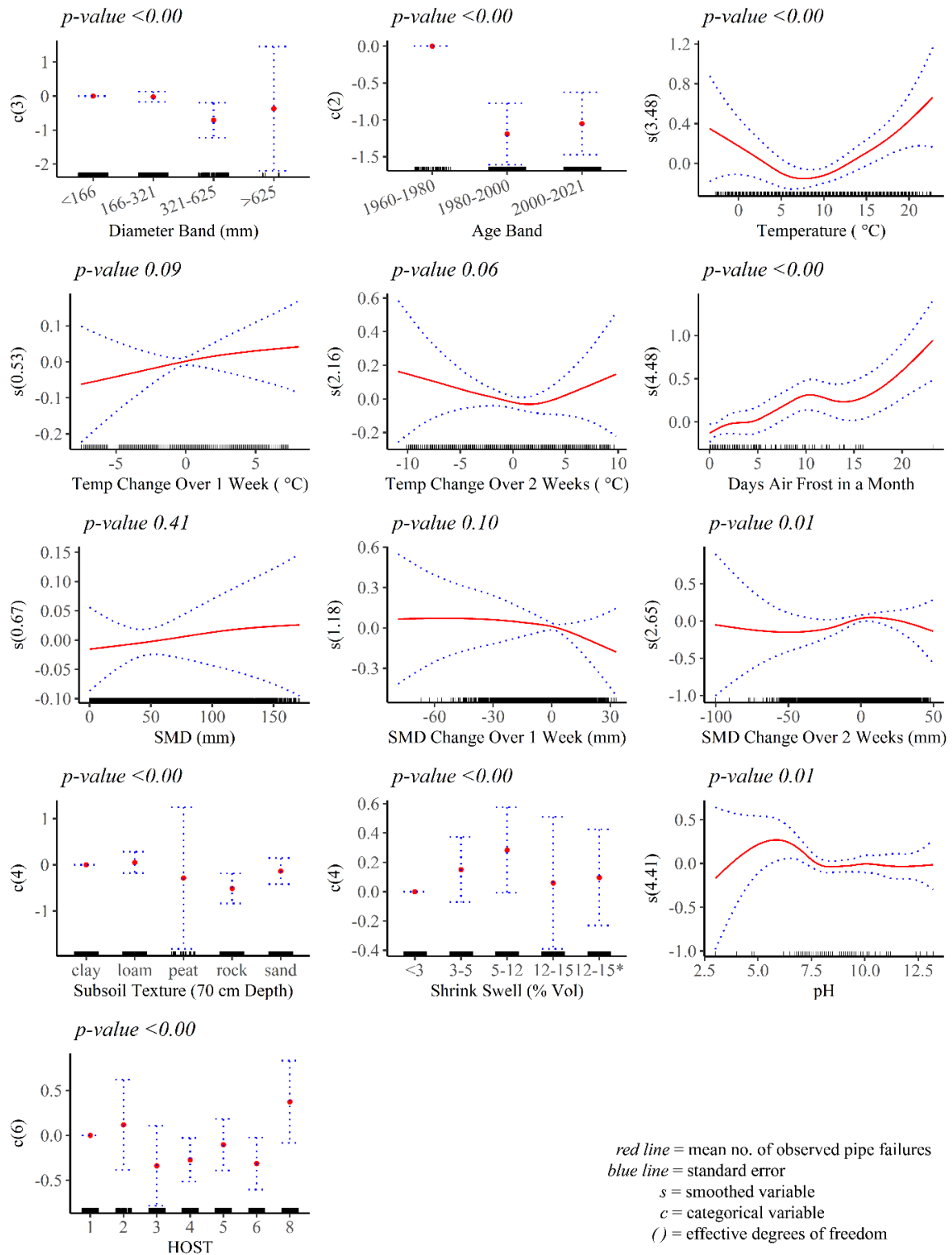


Figure 4-5 Generalized additive model - partial dependence plots showing the partial dependence of selected variables on Polyethylene pipe failures.

Figure 4-5 shows the GAM PDPs for PE and reveals that temperature and days air frost have the most effect on the number of PE pipe failures. Pipe failures increase for both low and high temperatures although confidence intervals increase for both high and low temperatures, while failures increase with more days air frost in a month up to 15 days, where the confidence intervals become high as the number of observations reduce. Temperature changes over one and two weeks, and SMD and SMD change over one and two weeks show no significance ($p\text{-value} > 0.01$), and there is no obvious variation observed in the number of PE pipe failures. Soil pH, whilst narrowly significant does not indicate any obvious pattern with PE pipe failure and alkaline or acidic soils. Diameter band also shows little variation between sizes but does suggest PE pipes of < 321 mm have the highest number of failures. Age band indicates older PE pipe installed between 1960 and 1981 are associated with a higher number of failures. Considering soil characteristics, soil shrink swell potential and subsoil type revealed little variation and therefore only a small effect on the number of pipe failures, while hydrologically impermeable (soft) soils (HOST category 8) comprising cover loam, clay with flints or plateau drift and loamy drift showed a higher number of pipe failures.

4.4 Discussion

4.4.1 Iron

Temperature, days air frost, SMD, peat and clay soils and pipes with a diameter < 166 mm are the most influential covariates on pipe failures. In interpreting the results, iron pipe failures increase during winter when the temperature is cold, being particularly vulnerable to temperatures below 3°C . This is attributed to the effects of air frost, and the subsequent cooling of soils adding additional stress on pipes (Fuchs-Hanusch et al., 2013; Kakoudakis et al., 2018; Wols et al., 2019). Cooling of soils is also exacerbated by the presence of water and therefore, has

more effect on soils with high water retention capacity. This is reflected in these data where the highest number of pipe failures for iron is observed in clay and peat soils, in line with previous studies (Wols et al., 2019). Peat and clay soils typically retain more moisture after rainfall due to small pores and a slow permeability of $< 0.1\text{mday}^{-1}$. Therefore, higher rainfall during October and November is retained by these soils remaining saturated during the winter. As soils take on water so the thermal conductivity increases, ceasing when the soil is saturated (the more water in the soil the colder the soils can become). However, when moisture turns to ice during low temperatures and prolonged air frost, the thermal conductivity increases once again allowing soils to become colder (Becker et al., 1992). Further to this soil cooling, the water retained in the soil pores freezes and increases in volume by approximately 9 % resulting in soil movement (Trickey et al., 2016). Frost fronts can also develop as a result of prolonged cold temperatures and can cause soil-related ground movement. However, in the UK, where frost typically lasts less than two weeks, frost fronts develop near the soil surface, and as a result exert only limited consequence on pipes, which are typically buried at an average depth of 85 cm. Occasionally longer periods of frost and consequent frost penetration to greater depths can lead to a higher number of pipe failures (Farewell et al., 2012a). The higher number of iron pipe failures observed in small diameter pipes $< 166\text{ mm}$, can be attributed to the pipe walls being thinner, meaning they have less resilience to ground movement and corrosion and are therefore more likely to fail. Joints in small diameter pipes have also been reported to be less reliable than in larger diameter pipes (Gould et al., 2011). There are also a significantly higher number of small diameter pipes compared to other pipe diameter sizes in the network.

4.4.2 Asbestos Cement

AC pipe failures are higher when temperature and SMD is high, being typically associated with the summer months. This is due to the effects of ground movement of highly expandable clay soils, as a result of associated clay minerals (i.e. smectite, montmorillonite and vermiculite) (Pritchard et al., 2015) which are

highly responsive to soil water content (O. G. Pritchard et al., 2015b); similar to findings in the Netherlands and Australia (Gould et al., 2011; Wols et al., 2019). A high number of AC pipe failures is especially seen in alluvial clay and peat soils that have a high shrink swell potential of 12–15 % volumetric expansion or greater if drained to a 2m depth. These types of conditions would require a high evapotranspiration rate to realise the potential shrink swell hazard which might be associated with a period of drought. In addition to soil shrink swell potential in clay soils, it is possible that pipe failures are also associated with concrete corrosion, since pipe failures are slightly higher in highly alkaline soils ($\text{pH} > 8.5$), although the effect is small. Other proxies for soil corrosion could potentially be used to determine the effects of corrosion on AC pipes. A small increase in pipe failures is associated with low temperatures between 5 °C and 0 °C, coupled with a near linear increase in pipe failures with an increase in days air frost in a month. This again highlights the influence of prolonged periods of frost on pipe failure as found in comparable literature that reported higher joint failures during winter for both AC and PVC pipes (Wols and Van Thienen, 2014). The effects of pipe age suggest a peak in pipe failures between 1940 and 1960. Pipe age is significant for AC and fewer failures are expected in new pipes. AC pipe was seen as a modern material during the 1940–1960s and was prolifically used. During this period 81 % of AC pipes in the network were installed.

4.4.3 Steel and Ductile Iron

For SDI pipes, the length of pipe in the network and number of pipe failures is considerably lower than other materials due to the material's cost. Days air frost and diameter size showed the strongest partial effects on pipe failures and suggests that SDI pipes fail as a result of the influence of frost on ground conditions, this has been discussed in detail for iron pipes (see Section 4.4.1). SDI pipe failures are highest during the winter months, so the results here are not surprising. SDI pipe show a small increase in pipe failures during the summer months (Barton et al., 2019). Observing the results to understand this trend, we look to see if soil shrink swell could be responsible. Despite the highest number

of SDI pipe failures being observed in clay soil, there is little evidence to suggest soil shrink swell is the cause, since SDI pipe failures reduce as the potential for volumetric expansion increases. Therefore, other factors not included in this study may explain this seasonal trend observed, such as water usage or pipe pressure.

4.4.4 Polyvinyl Chloride

The PDPs suggest that a limited number of environmental variables effect PVC pipe failures. PVC failures do increase as temperature increases, with the highest mean failure rate observed during the hottest temperatures, similar to AC pipes (Wols and Van Thienen, 2014); however above 16 °C the uncertainty bounds of these data also increase. PVC shows no relationship with soil shrink swell, with little variation in pipe failures from low to high classes and is likely to be related to the plasticity of the material and the flexibility of the push fit joint most commonly used in PVC pipes. PVC pipe failures were highest in peat subsoils (noting however the uncertainty bounds of peat were very high) and in loam soil, but the effects of soil are small. It is also worth noting that flexible fitting joints such as the push fit joints typically used on PVC pipes can absorb soil displacement to a certain degree, which reduce stress on PVC pipes in soils acceptable to ground movement (Wols and Van Thienen, 2014). Considering this and the small effects of SMD up to approximately 120 mm, we suggest a small effect from drying soils, but this does not explain the high summer failures for PVC which are likely to be associated with other contributing factors. Previous studies have suggested that internal pressure from increased water demand during the summer months may be the reason for PVC pipe failures (Wols et al., 2019). This is reported as particularly detrimental in push-fit joints since they do not provide restraint from thrust from internal pressure forces, resulting in axial pull out or disconnection (Arsénio et al., 2013; Sundberg, 2017). In these data we did not have access to either pressure data or failure type, therefore further investigation is required to support the literature in this area and understand the main failure mechanism. Failures may also be associated with soil washout

potential, where an existing pipe failure can wash soil from under another, causing the pipe to bridge, sag and eventually fail (Pritchard et al., 2013). Considering the effects of days air frost, PVC shows a degree of vulnerability to long periods of frost, similar to iron pipes and as reported by existing literature (Wols and Van Thienen, 2014). However, considering the uncertainty interval the effect may be minor.

4.4.5 Polyethylene

PE pipes show little variation in number of failures in the PDPs. Temperature shows a higher number of failures at low temperatures below 5 °C, which is likely to be related to the effects of frost on soils since failures increase as the number of days air frost increases. However, the wide confidence intervals suggest the effects are small. The winter failures may be associated with joint failure as per PVC and AC (Ruchti, 2017) since the material's ability to withstand thermal expansion and contraction (plasticity) makes it more resilient to ground movement. However, this could only occur on joints that were not properly fused together, since PE pipes connected correctly through electrofusion rarely fail. Due to the plasticity of PE pipes and limited partial effects in different soils and shrink swell, the increase in pipe failures during high temperatures are likely to be related to other factors such as increased water demand. Age of PE pipe shows that early pipes installed between 1960 and 1981 are associated with a higher number of failures (Figure 4-5). This could be due to early PE material being a different grade of PE or improvements in joining pipes shifting from plate welding and clamping to electrofusion (Barton et al., 2019).

4.5 Concluding Remarks

Using a multivariate GAM with PDPs provided a flexible and useful tool for revealing insight into complex relationships and structure in these data. The findings of this study contribute to existing literature and can be used to aid a further understanding of environmental interactions and how they affect pipe failures for different material types. A natural extension of these findings would

be to build future statistical models with knowledge of which environmental variables are likely to explain pipe failures for each material type. The results found here for an entire UK network are similar to comparable studies from other countries (Netherlands (Wols et al., 2019; Wols and Van Thienen, 2014) Austria (Fuchs-Hanusch et al., 2013) and Australia (Gould et al., 2011), and will be useful to water utility companies, who could use the information to understand failures in similar network conditions. The prominent findings of the study include:

- Some of the continuous variables smoothed by the GAM displayed low effective degrees of freedom. Where this occurred, the variables may provide better correlation as linear predictors.
- Diameter band is important for all materials, except PVC. Higher failure rates occurred in pipes with a diameter of < 166 mm.
- Pipe age is important for all pipe materials except iron and PVC.
- Severe temperatures have a significant influence on pipe failures for all materials, however the influence of temperature strongly depends on the type of pipe material.
- Soil pH, considered a marker for corrosion, has far less influence than was initially anticipated. Other soil corrosion proxies such as a soil corrosion index may yield better results.
- A rapid change in both temperature and SMD as determined in this study has a weak effect on pipe failures for all materials.
- Iron and SDI pipes fail mainly in the winter, during temperatures < 5 °C and long periods of air frost. Clay and peat soils with high moisture retention show the highest number of failures during the winter.
- AC pipes mainly fail during the summer as a result of increasing temperatures, high SMD and subsequent shrink swell in highly shrinkable clay and peat dominant soils causing soil movement.
- PVC pipe failures increase during the summer and are affected by temperature and SMD. However, ground movement does not appear to

be the only cause. Therefore, other factors not explored in this study must be able to explain PVC summer failures.

- PE pipes exhibit a higher number of failures in severely cold temperatures with prolonged air frost.

Understanding the effects of different environmental factors on pipe failure is an important step when building data-driven, statistical infrastructure models. This understanding can help to reduce common confusion of correlation and causation effects and can help in the selection of appropriate variables and their derivatives. Such information can lead to more accurate and robust models, which in turn can be used by utilities to make informed network management decisions, aiming at reducing water lost through network leaks and helping facilitate future water demand.

4.6 Author Contributions

N.B.: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing—Original Draft, Visualization, Data Curation. T.F.: Project Administration, Validation, Formal Analysis, Resources, Writing—Review and Editing, Supervision, Data Curation, Funding Acquisition. S.H.: Writing—Review and Editing, Supervision, Funding Acquisition.

4.7 References

- Arsénio, A.M., Pieterse, I., Vreeburg, J., de Bont, R., Rietveld, L., 2013. Failure mechanisms and condition assessment of PVC push-fit joints in drinking water networks. *J. Water Supply Res. Technol.* 62, 78–85. <https://doi.org/10.2166/aqua.2013.026>
- Barton, N.A., Farewell, T.S., Hallett, S.H., Acland, T.F., 2019. Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Res.* 164, 114926. <https://doi.org/10.1016/j.watres.2019.114926>

- Becker, B.R., Misra, A., Fricke, B.A., 1992. Development of correlations for soil thermal conductivity. *Int. Commun. Heat Mass Transf.* 19, 59–68. [https://doi.org/10.1016/0735-1933\(92\)90064-O](https://doi.org/10.1016/0735-1933(92)90064-O)
- Boorman, D.B., Hollis, J.M., Lilly, A., 1995. Hydrology of soil types: a hydrologically based classification of the soils of United Kingdom, Institute of Hydrology.
- Bruaset, S., Sægrov, S., 2018. An analysis of the potential impact of climate change on the structural reliability of drinking water pipes in cold climate regions. *Water (Switzerland)* 10. <https://doi.org/10.3390/w10040411>
- Buja, A., Hastie, T., Tibshirani, R., 1989. Linear Smoothers and Additive Models. *Ann. Stat.* 17, 453–510.
- Clark, R.M., Stafford, C.L., Goodrich, J.A., 1982. Water distribution systems: a spatial and cost evaluation. *J. Water Resour. Plan. Manag.* 108, 243–256.
- De Oliveira, D.P., Neill, D.B., Garrett, J.H., Soibelman, L., 2011. Detection of patterns in water distribution pipe breakage using spatial scan statistics for point events in a physical network. *J. Comput. Civ. Eng.* 25, 21–30. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000079](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000079)
- Dingus, M., Haven, J., Austin, R., AWWA Research Foundation., 2002. Non-destructive, non-invasive assessment of underground pipelines. AWWA Research Foundation and American Water Works Association.
- Environment Agency, 2018. Environment Agency calls for action on water efficiency - GOV.UK [WWW Document]. URL <https://www.gov.uk/government/news/environment-agency-calls-for-action-on-water-efficiency> (accessed 10.30.19).
- ESRI, 2019. ArcGIS Online | Interactive Maps Connecting People, Locations & Data [WWW Document]. URL <https://www.arcgis.com/index.html> (accessed 2.6.19).

- Farewell, T.S., Hallett, S.H., Hannam, J.A., Jones, R.J.A., 2012a. Soil impacts on national infrastructure in the UK, Infrastructure Transitions Research Consortium.
- Farewell, T.S., Hallett, S.H., Truckell, I.G., 2012. Soil and climatic causes of water mains infrastructure bursts. NSRI Res. report. NSRI, Cranf. Univ. UK. 111pp.
- Farewell, T.S., Jude, S., Pritchard, O., 2018. How the impacts of burst water mains are influenced by soil sand content. Nat. Hazards Earth Syst. Sci. 18, 2951–2968. <https://doi.org/10.5194/nhess-18-2951-2018>.
- Farrow, J., Jesson, D., Mulheron, M., Nensi, T., Smith, P., 2017. Achieving zero leakage by 2050: the basic mechanisms of bursts and leakage, UK Water Industry Research Limited. London.
- Fasiolo, M., Nedellec, R., Goude, Y., Wood, S.N., 2019. Scalable Visualization Methods for Modern Generalized Additive Models. J. Comput. Graph. Stat. 1–9. <https://doi.org/10.1080/10618600.2019.1629942>
- Fuchs-Hanusch, D., Friedl, F., Scheucher, R., Kogseder, B., Muschalla, D., 2013. Effect of seasonal climatic variance on water main failure frequencies in moderate climate regions. Water Supply 13, 435–446. <https://doi.org/10.2166/ws.2013.033>
- Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation. J. Comput. Graph. Stat. 24, 44–65. <https://doi.org/10.1080/10618600.2014.907095>
- Gould, S.J.F., Boulaire, F.A., Burn, S., Zhao, X.L., Kodikara, J.K., 2011. Seasonal factors influencing the failure of buried water reticulation pipes. Water Sci. Technol. 63, 2692–2699. <https://doi.org/10.2166/wst.2011.507>

- Goulter, I.C., Kazemi, A., 1988. Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Can. J. Civ. Eng.* 15, 91–97. <https://doi.org/10.1139/l88-010>
- Hallett, S.H., Sakrabani, R., Keay, C.A., Hannam, J.A., 2017. Developments in land information systems: examples demonstrating land resource management capabilities and options. *Soil Use Manag.* 33, 514–529. <https://doi.org/10.1111/sum.12380>
- Hastie, T., Tibshirani, R., 1986. Generalized additive models. *Stat. Sci.* 1, 297–310. <https://doi.org/10.1214/ss/1177013604>
- He, S., 2004. Generalized Additive Models for Data With Concurvity: Statistical Issues and a Novel Model Fitting Approach. University of Pittsburgh.
- Heinze, G., Wallisch, C., Dunkler, D., 2018. Variable selection – A review and recommendations for the practicing statistician. *Biometrical J.* 60, 431–449. <https://doi.org/10.1002/bimj.201700067>
- Hough, M.N., Jones, R.J.A., 1997. The United Kingdom Meteorological Office rainfall and evaporation calculation system: MORECS version 2.0-an overview. *Hydrol. Earth Syst. Sci.* 1, 227–239. <https://doi.org/10.5194/hess-1-227-1997>
- Kakoudakis, K., Behzadian, K., Farmani, R., Butler, D., 2017. Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K-means clustering. *Urban Water J.* 14, 737–742. <https://doi.org/10.1080/1573062X.2016.1253755>
- Kakoudakis, K., Farmani, R., Butler, D., 2018. Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. *J. Hydroinformatics* 20, 1191–1200. <https://doi.org/10.2166/hydro.2018.152>

- Kirby, P.C., 1981. PVC Pipe Performance in Water Mains and Sewers, in: International Conference on Underground Plastic Pipe ASCE. New Orleans, LA, pp. 161–174.
- Marra, G., Wood, S.N., 2011. Practical variable selection for generalized additive models. *Comput. Stat. Data Anal.* 55, 2372–2387. <https://doi.org/10.1016/j.csda.2011.02.004>
- Met Office, 2019. East Anglia Days of Air Frost [WWW Document]. URL https://www.metoffice.gov.uk/pub/data/weather/uk/climate/datasets/AirFrost/ranked/East_Anglia.txt (accessed 2.6.19).
- Ortega-Arranz, H., Torres, Y., Llanos, D.R., Gonzalez-Escribano, A., 2014. The All-Pair Shortest-Path Problem in Shared-Memory Heterogeneous Systems, in: Jeannot, E., Zilinskas, J. (Eds.), *High-Performance Computing on Complex Environments*. John Wiley & Sons, Ltd, pp. 283–297.
- Pelletier, G., Mailhot, A., Villeneuve, J.P., 2003. Modeling Water Pipe Breaks—Three Case Studies. *J. Water Resour. Plan. Manag.* 129, 115–123. [https://doi.org/10.1061/\(ASCE\)0733-9496\(2003\)129:2\(115\)](https://doi.org/10.1061/(ASCE)0733-9496(2003)129:2(115))
- Pritchard, Oliver; Farwell, Timothy; Hallett, S., 2015. Soil-related geohazard assessment for climate-resilient UK infrastructure. *Cranf. Univ.* 0–400.
- Pritchard, O., Hallett, S.H., Farewell, T.S., 2013. Soil Corrosivity in the UK – Impacts on Critical Infrastructure. *Infrastructure Transitions Res. Consort.* 55.
- Pritchard, O.G., Hallett, S.H., Farewell, T.S., 2015. Probabilistic soil moisture projections to assess Great Britain's future clay-related subsidence hazard. *Clim. Change* 133, 635–650. <https://doi.org/10.1007/s10584-015-1486-z>

- Pritchard, O.G., Hallett, S.H., Farewell, T.S., 2014. Soil impacts on UK infrastructure: current and future climate. *Proc. Inst. Civ. Eng. - Eng. Sustain.* 167, 170–184. <https://doi.org/10.1680/ensu.13.00035>
- R Core Team, 2018. R: The R Project for Statistical Computing [WWW Document]. URL <https://www.r-project.org/> (accessed 2.6.19).
- Rajani, B., Kleiner, Y., Sink, J.-E., 2012. Exploration of the relationship between water main breaks and temperature covariates. *Urban Water J.* 9, 67–84. <https://doi.org/10.1080/1573062X.2011.630093>.
- Kleiner, Y., Rajani, B., 1999. Using limited data to assess future needs. *J. Am. Water Works Assoc.* 91, 47–61. <https://doi.org/10.1002/j.1551-8833.1999.tb08664.x>
- Ratner, B., 2010. Variable selection methods in regression: Ignorable problem, outing notable solution. *J. Targeting, Meas. Anal. Mark.* 18, 65–75. <https://doi.org/10.1057/jt.2009.26>
- 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. *Reliab. Eng. Syst. Saf.* 196, 106754. <https://doi.org/10.1016/j.res.2019.106754>
- Røstum, J., 2000. Statistical Modelling of Pipe Failures in Water Networks. Norwegian University of Science and Technology.
- Ruchti, G.F., 2017. Water Pipeline Condition Assessment. American Society of Civil Engineers (ASCE). <https://doi.org/10.1061/9780784414750>
- Scheidegger, A., Leitão, J.P., Scholten, L., 2015. Statistical failure models for water distribution pipes – A review from a unified perspective. *Water Res.* 83, 237–247. <https://doi.org/10.1016/j.watres.2015.06.027>
- Sundberg, C., 2017. Fundamentals of Differential Settlement of Pipelines, in: Pridmore, A.B., Geisbush, J. (Eds.), *Sessions of the Pipelines Conference*,

- Phoenix, AZ, AUG 06-09, 2017. Pipelines 2017. Planning and design: proceedings of sessions of the Pipelines 2017 Conference, August 6-9, 2017, Phoenix, Arizona, pp. 469–481.
- Sunita, Garg, D., 2018. Dynamizing Dijkstra: A solution to dynamic shortest path problem through retroactive priority queue. J. King Saud Univ. - Comput. Inf. Sci. <https://doi.org/10.1016/j.jksuci.2018.03.003>
- Trickey, S.A., Moore, I.D., Balkaya, M., 2016. Parametric study of frost-induced bending moments in buried cast iron water pipes. Tunn. Undergr. Sp. Technol. 51, 291–300. <https://doi.org/10.1016/j.tust.2015.10.028>
- Wiley, M., Wiley, J.F., 2019. GAMs, in: Advanced R Statistical Programming and Data Models. Apress, Berkeley, CA, pp. 165–224. https://doi.org/10.1007/978-1-4842-2872-2_5
- Wols, B. A., Van Thienen, P., 2014. Modelling the effect of climate change induced soil settling on drinking water distribution pipes. Comput. Geotech. 55, 240–247. <https://doi.org/10.1016/j.compgeo.2013.09.003>
- Wols, B.A., Van Thienen, P., 2014. Impact of weather conditions on pipe failure: a statistical analysis. J. Water Supply Res. Technol. 63, 212–223. <https://doi.org/10.2166/aqua.2013.088>
- Wols, B.A., Vogelaar, A., Moerman, A., Raterman, B., 2019. Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. Water Supply 19, 404–416. <https://doi.org/10.2166/ws.2018.085>
- Wood, S.N., 2006. Generalized additive models: An introduction with R, First Edit. ed. Chapman and Hall/CRC.
- Wood, S.N., Augustin, N.H., 2002. GAMs with integrated model selection using penalized regression splines and applications to environmental modelling. Ecol. Modell. 157, 157–177. [https://doi.org/10.1016/S0304-3800\(02\)00193-X](https://doi.org/10.1016/S0304-3800(02)00193-X)

Wood, S.N., Goude, Y., Shaw, S., 2015. Generalized additive models for large data sets. J. R. Stat. Soc. Ser. C Appl. Stat. 64, 139–155.

<https://doi.org/10.1111/rssc.12068>

Zhou, Y., 2018. Deterioration and Optimal Rehabilitation Modelling for Urban Water Distribution Systems, Deterioration and Optimal Rehabilitation Modelling for Urban Water Distribution Systems.

<https://doi.org/10.1201/9780429451799>

5 An Evolution of Statistical Pipe Failure Models for Drinking Water Networks: A Targeted Review

Objective 3	Chapter 5: An evolution of statistical pipe failure models: a targeted review	<ul style="list-style-type: none">▪ Review existing literature on statistical pipe failure models.
		<ul style="list-style-type: none">▪ Identify the most appropriate model, response type, prediction interval and spatio-temporal scale given the available variables identified in Objective 2.

The following chapter provides a summary of statistical pipe failure models classified into deterministic, probabilistic and machine learning, in a targeted review and fulfils Objective 3. The chapter is taken directly, without further edits, from the following article published in International Water Associations Journal, Water Supply, in line with the Creative Commons Attribution 4.0 international license agreement:

Barton, N.A., Hallett, S.H., Jude, S.R., 2021. An Evolution of Statistical Pipe Failure Models for Drinking Water Networks: A Targeted Review. Water Supply. <https://doi.org/10.2166/ws.2022.019>

An Evolution of Statistical Pipe Failure Models for Drinking Water Networks: A Targeted Review

Neal Andrew Barton^a, Stephen Henry Hallett^{a*}, Simon Richard Jude^a and Trung Hieu Tran^b

^a School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

^b Centre for Competitive Creative Design (C4D), School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

**Corresponding author: Email: s.hallett@cranfield.ac.uk; Tel: +441234752978*

Abstract

The use of statistical models to predict pipe failures has become an important tool for proactive management of drinking water networks. This targeted review provides an overview of the evolution of existing statistical models, grouped into three categories: deterministic, probabilistic and machine learning. The main advantage of deterministic models is simplicity and relative minimal data requirement. Deterministic models predicting failure rates for the network or large groups of pipes perform well and are useful for shorter prediction intervals that describe the influences of seasonality. Probabilistic models can accommodate randomness and are useful for predicting time to failure, interarrival times and the probability of failure. Probability models are useful for individual pipe models. Generally, machine learning describes large complex data more accurately and

can improve predictions for individual pipe failure models yet are complex and require expert knowledge. Non-parametric models are better suited to the non-linear relationships between pipe failure variables. Census data and socio-economic data requires further research. The complexity of choosing the most appropriate statistical model requires careful consideration of the type of variables, prediction interval, spatial level, response type and level of inference is required.

Key Words: Drinking water; data analytics; machine learning; pipe failure; statistical modelling.

Highlights:

- Discusses key statistical models, including regression, probability, and machine learning.
- Reviews fundamental characteristics, limitations, and progress.
- Compares between the main outcomes and discusses future research.
- Synthesises the findings and includes an aid for different decision-making contexts.

5.1 Introduction

Developing reliable infrastructure decision-making tools is essential for managing large deteriorating Water Distribution Networks (WDNs), which pose economic, societal, and environmental threats if they fail (St. Clair and Sinha, 2012). Water companies are interested in proactive management to determine which assets are likely to fail in the short (monthly, inter-annual), medium (annual) or long term (asset management period) (Watson et al., 2004). One area of innovation and opportunity is predictive pipe failure modelling, an area that focuses on predicting the time and location of pipe failures (also referred to as breaks), and a variety of such statistical models exist in the literature. Early pipe failure models were developed with the emergence of modern asset management and a drive to

understand the cost of pipe failures. Over the past few decades, many models have been developed, benefitting from advances in computing and geographical information systems and databases able to collect and manipulate the substantive volumes and variety of data involved (Deadman, 2010). Water companies now use pipe failure models in seeking new ways to understand and manage assets. Many such approaches focus on medium (annual) to long term (> annual) planning and concern the vital area of asset management, assessing the condition of a WDN and exploring strategies for pipe replacement, rehabilitation, and maintenance. However, few approaches focus on the short term, which would provide a framework to help operational management, an area of continued discussion. Model development is based on an initial assessment of the desired output, the specific needs of each water company and the performance required across varying geographic locations. Other key variables concern the age of the WDN, the mix of material types, and the range of variables available within the data. It is essential to understand the different modelling approaches, each with its own merits and reasons for achieving different outcomes.

Pipe failure models are broadly classified as heuristic, physical or statistical. Several comprehensive reviews summarise pipe failure models extensively (Kleiner and Rajani, 2001; Nishiyama and Fillion, 2013; Scheidegger et al., 2015; St. Clair and Sinha, 2012; Wilson et al., 2017). These literature reviews are systematic and focus on a particular period (the last decade) or type of model (deterministic and probabilistic), or specific aspects such as large diameter pipes. These review papers do not capture the current machine learning models widely used today. The idea of this exposition is to look purposefully at the progression of pipe failure models encompassing a range of statistical models, including deterministic, probabilistic and machine learning used in literature. We use a targeted or focussed review process, selecting high-quality articles over time that help identify the trend and current state of pipe failure models, rather than an exhaustive list of literature or literature from a specific period or type. This

approach is a means of discussing a narrative of change and progression in statistical pipe failure models. Key historical models are discussed, as pipe failure models evolved from simple deterministic models to current machine learning, describing some of the fundamental issues and discussing limitations and progress. The need to detail machine learning is essential given the increasing trend towards these models in the past decade. Therefore, an emphasis is given on various methods being leveraged for their capability to describe and encapsulate complex relationships and return more accurate predictions. Regulatory incentives driving innovation (Ofwat, 2020c) are likely to spur further desire for machine learning and data analytics utilised to support data management of these substantive data repositories (Ponce Romero et al., 2017). Figure 5-1 presents a taxonomy of the different model types discussed in this review.

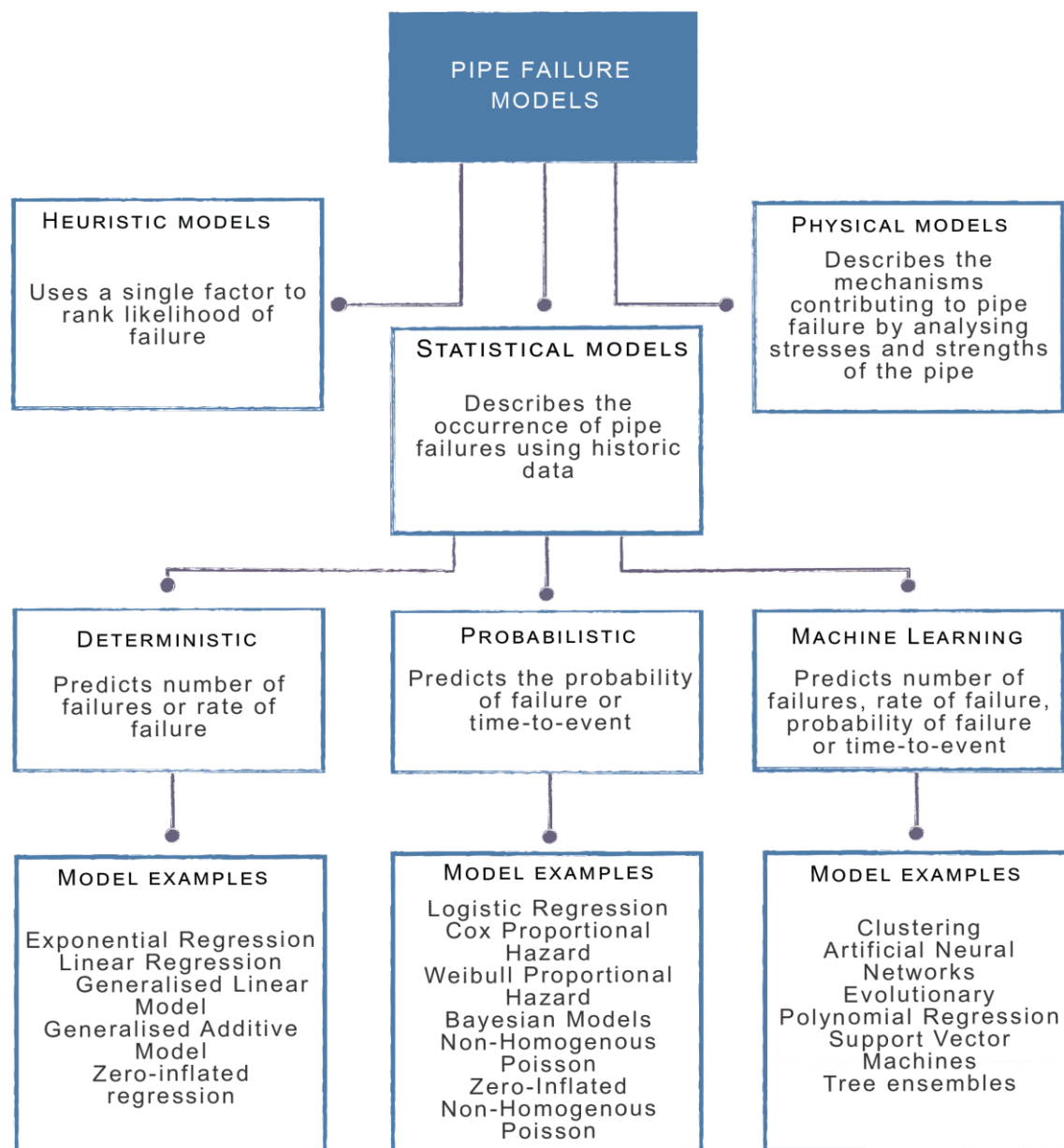


Figure 5-1 Taxonomy of Pipe Failure Models (adopted from Snider and McBean (2020)).

5.2 Heuristic Models

Water utilities have historically adopted heuristic models to calculate the failure rates of pipes to prioritize replacement scheduling. Heuristic models are often not included in reviews; however, they represent a common approach for assessing infrastructure where factors and mechanisms are not well understood. Instead,

Heuristic models use descriptive analytics such as assumptions, expert opinion, or subjective weightings on leakage rates, material type, and age to determine the necessity of a pipe's replacement. Heuristic models are accessible to water companies and do not require extensive prior knowledge of variable selection and data pre-processing. Furthermore, limited or missing variables will not affect overall accuracy. However, heuristic models are rarely optimal since they rely on subjective opinions and fail to capture the potential risks or accurately describe future failure rates (St. Clair and Sinha, 2012; Fitchett et al., 2020; Snider and McBean, 2020).

5.3 Physical Models

Physical models (mechanistic models) ordinarily aim to describe the mechanisms that contribute to pipe failure by analysing the pipe's stresses and strengths and have been reviewed in detail by Rajani and Kleiner (2001). Relating these mechanisms involves analysing detailed data on pipe structural properties (quality of installation, material type, age, diameter, and manufacturing details), internal and external loads, proximal factors (environmental conditions), and material deterioration (Rajani and Kleiner, 2001). Physical models do not require large quantities of historical data for calibration or operation (Wilson et al., 2017); alternatively, pipe investigation is necessary to determine the pipe's state of deterioration and site investigations can be expensive and time consuming. Once created, the model represents the specific pipe and does not extrapolate to other areas of the WDN, limiting the use of the model. For these reasons, physical models are ordinarily limited to large supply mains (Wilson et al., 2017).

5.4 Statistical Models

Statistical models aim to predict pipe failures from historical data containing several correlated variables, which may be categorised broadly as pipe-intrinsic, environmental and operational (Barton et al., 2019, 2020). Finding an appropriate model assumes that the response is not random but is functionally related to the variables. Variables take the form $x_1, x_2, \dots, x_n$, and form a n -dimensional input

dataset typically denoted as X . In pipe datasets, ordinarily X is represented by quantitative data, being either inherently categorical (soil or pipe material) or continuous (e.g., pH). These variables remain mainly static whilst other factors include a dynamic time dependent aspect (such as with weather and pipe pressure (Wols et al., 2019)). The response variable is similarly defined by an output space of Y described from $y_1, y_2, \dots, y_m$, either numerical, in which case it is regression and defined by the function $\varphi : x \rightarrow y$, where $y = \mathbb{R}$, here φ represents the prediction, categorical and therefore classification where $y = c_1, c_2, \dots, c_m$ a set of classes, or probability where $y = \mathbb{P}$ the probability of occurrence (Louppe, 2014). Statistical models are favoured over physical models because they are cost and time effective, do not require information on the physical pipe mechanisms (such as deterioration) and are easily extrapolated to other pipes with similar characteristics across an entire WDN (Kleiner and Rajani, 2001). Statistical models are classified in numerous ways, but commonly into deterministic, probabilistic and machine learning (Nishiyama and Fillion, 2014).

5.4.1 Deterministic Models

5.4.1.1 Single-Variate Regression

Linear models are simple deterministic models that have been used widely in statistics for many decades. The single-variate linear model represents a one-dimensional space with a linear function given as:

$$y_i = \beta_0 + \sum_i \beta_i x_i + \varepsilon \quad \text{Eq.5-1}$$

where β_0 is the intercept, β_i is the weight of x_i and ε the error term. Its simplicity delivers reliability and interpretability, which is advantageous to water companies (sometimes referred to as a white-box approach). Models using this approach have been used to return the number of failures or by dividing the number of failures by pipe length as a natural estimator of failure rate. Early aggregate models predicted the number of failures as a function of time, using linear relationships. Kettler and Goulter (1985) developed a time-linear regression

model using pipe age and diameter, given by $N = k_0 A$, where N is the number of failures, k_0 is the regression parameter, and A is age of pipe at first failure. Kettler and Goulters (1985) model was applied to a small WDN in Winnipeg, Manitoba, using a ten-year failure record, and revealed a strong negative correlation between failure rates and diameter size. Despite the term of linear, linear models can also model and represent non-linear relationships. Shamir and Howard (1979) developed one of the first models on a small WDN of pipes, a single-variate time-exponential model using pipe age to predict the number of failures per year per 1000 ft of pipe, such that: $N(t) = N(t_0) \exp^{A(t-t_0)}$, where t is time, $N(t)$ is the rate of failure at time t , t_0 is the reference year, and A is a failure rate growth coefficient. These early studies were limited by the single-variate approach, which proved of limited practicality for asset management due to its scale of operation, and for determining relationships with other important variables. Shamir and Howard (1979) also used a pipe group approach, assuming similar failure rates for all pipes within the group, which is not often the case.

5.4.1.2 Multivariate Regression

Studies began to focus on predicting the number of failures as a function of multiple factors (or covariates) (Andreou et al., 1987a). Kleiner and Rajani (2000) were the first to include dynamic time-related weather variables (soil moisture, temperature, and frost) using multi-variate exponential models, which improved predictive capability and described the influence of the weather on failure rates. The main limitation in the study was the removal of static variables, which are essential since static variables such as soil are present in all WDNs and are known to influence pipe breakage (B Rajani et al., 1996). Kleiner and Rajani (2002) again used the multi-variate exponential model and the same weather variables on groups of pipes, this time noting that more variables led to improved accuracy in the predictions. Additionally, operational factors such as cathodic protection were used, as to was mains replacement rate, not used before. These variables were able to accommodate data bias in the WDN dataset

caused by aggressive pipe replacement and the removal of several high failing pipes and essential information. The use of both environmental and operational variables meant the model could account more accurately for ageing over time, yet both models were simplistic in their approach and showed only moderate accuracy overall due to many conditional assumptions.

To improve multivariate analysis, different models were explored and further operational factors such as pipe pressure were introduced. Wang et al. (2009) predicted annual failure rates for individual pipes using pipe-intrinsic factors for five pipe materials using multiple regression. The study was completed on a 432 km WDN in St Foy, Quebec City, based on data collected between 1987 and 2001. The approach showed a good level of accuracy through best-subset regression, with pipe length being an important factor; R^2 of 68.9, 65.0, 71.5, 78.4, and 81.3 for cast iron, ductile iron (without lining), ductile iron (with lining), PVC, and hyprescon pipes, respectively. The major limitation is the focus on how annual failure rates affected variables, suggesting that longer pipes had lower failure frequency and that annual failure rates increased with age but did not predict when the subsequent failure would occur on a pipe. In a comparison of different models, Asnaashari et al. (2009) modelled failures for groups of pipes on a small-sized WDN (554 pipes) for ten years of pipe failure records, comparing multiple regression to Poisson regression for four materials separately (AC, CI, DI and PE). The authors reported a higher accuracy in the Poisson models compared to the multiple regression models; R^2 between 0.71 and 0.95, and between 0.52 and 0.88, respectively. The Poisson model was said to perform well because it can handle non linear relationships more accurately. However, a deviation and Pearson chi-square test calculated over-dispersion, showing low unity between the metrics, signifying over-dispersion due to zero-inflated data. The study did not include dynamic time-related variables, which was another limitation but focussed on intrinsic pipe factors of wall thickness, pipe diameter, pipe length, cover depth, location, failure history and pipe pressure

5.4.1.3 Generalised Models

Moving towards addressing the pipe grouping issue, failure models started to predict individual pipes. The Generalised Linear Model (GLM) was used because of its capability to model complex relationships through a differentiable monotonic link function (exponential, Poisson or logistic are favoured), allowing for non-normal distributions and taking the general form (Yamijala et al., 2009):

$$y_i = \beta_o + b_i x_i \quad \text{Eq.5-2}$$

Yamijala et al. (2009) compared multiple models on a medium-sized WDN in Texas (<1,600 km) between 2000 to 2005, including exponential regression, Poisson GLM, and logistic regression, including important factors such as pressure, soil parameters and temperature. The model predicted failures for individual pipes on a six-month interval to capture temperature variation. The authors reported no good results due to data imbalance between the number of failures and non-failures observed in the data, potentially imbalanced further by the short prediction period. Nevertheless, the authors suggest that logistic GLM provided valid test model estimates of pipe reliability at this spatiotemporal scale since predicting the probability of a discrete outcome is often helpful for water companies (Yamijala et al., 2009). The linear, exponential and Poisson models were performed on a reduced data set, consisting of only pipes with failures to avoid the zero-inflation problem. The linear and exponential models were fit with fewer variables than the Poisson and logistic model, which also used an iterative process to drop variables for a parsimonious model. In this instance, the Pearson's correlation coefficient was low for the Poisson model, which suggested overdispersion was not a problem in the data, unlike that found by Asnaashari et al. (2009), potentially due to using only pipes with failures. In a similar study by Motiee and Ghasemnejad (2019), the authors compared four models on individual pipes for a medium-sized WDN (583 km) in Tehran over four years of failure records (2004-2007), including Linear regression, Exponential regression, Poisson GLM, and Logistic GLM. The authors found similar results, with the Logistic regression model observed to be the most useful, but the study failed to

include a range of important variables such as soil, temperature or pipe pressure and did not handle the zero-inflated data.

5.4.1.4 Zero-Inflated Models

Other studies have attempted to use models to accommodate the zero-inflated data inherent in most pipe failure datasets. The Zero Inflation Poisson (ZIP) model handles zero-inflation by allowing different probabilities through two components; the first generates zero counts, and the second generates counts with probability, some of which may be zero. In a study by Konstantinou and Stoianov (2020), several models were compared and evaluated on a network of 374 km between 2003-2016, using 550 pipe failures. A ZIP GLM and negative binomial GLM were amongst those used but were the only models that could manage the zero-inflated dataset for individual pipes. Both models poorly fitted the data and underestimated failures in the network; R^2 of 0.091 and 0.082, respectively. The authors recognised a lack of complete (no environmental variables) and reliable data, which is likely to have reduced the accuracy of the model, yet the machine learning models were capable of accurate predictions. Therefore, the zero-inflated and negative binomial models did not perform well on this dataset, but no explanation was given. The models were applied to pipe segments in two scenarios, first for each pipe and second for pipe lengths partitioned into 200 m lengths. The model performed better on individual pipe lengths because of the direct classification of the pipe type in the first scenario. The variables provided were sufficient and were standardised with a mean of zero and standard deviation of one, but with few pipe failures in the dataset, groups of pipes may have yielded better results.

5.4.1.5 Comments

The simplistic mathematical framework of deterministic models has made them popular with water companies, especially given the ease of interpreting their results. Single variate models such as the time linear and time exponential models are easy to develop, and early models that predicted the whole network condition were informative. However, these approaches were of limited use since

they do not provide enough information to describe pipe failures in detail, and the parametric design fails to describe the non-parametric nature of pipe failures. The introduction of multivariate models improves predictions by providing further information, and multivariate regression and Poisson regression models were used to some success when predicting failures and deriving further understanding. The introduction of dynamic variables showcased the influence of seasonal weather variation but can only be used with relatively large groups of pipes or the entire network and is not suitable for individual pipes. Further introduction of operational variables was necessary, yet studies observed that data was not always available. Predicting the number of failures or failure rate using pipe length as a normalisation factor was well suited to these high scale models since failure frequencies are discrete and non-negative (Asnaashari et al., 2009).

A practical step in pipe failure models was the use of multivariate models predicting individual pipes, due to the apparent advantages to water companies. However, this proved mathematically problematic due to the zero-inflation problem caused by an insufficient number of failures for each pipe. The GLM model extended more basic multivariate regression by allowing greater flexibility via a link function, with the Poisson distribution proving more successful than others, such as the exponential or linear distributions. One assumption of the Poisson model is that the response variance is equal to the mean, yet this is seldom the case in many pipe failure data, especially for zero-inflated data, resulting in over dispersion (Asnaashari et al., 2009). When this assumption is violated, the methodology does not meet the Poisson distribution assumptions (Asnaashari et al., 2009; Kleiner and Rajani, 2012) and overestimate pipe failures with residuals showing significant error and bias (Konstantinou and Stoianov, 2020), and other studies suggesting no useful results (Yamijala et al., 2009). By definition, it can be suggested that predicting failure rates is problematic for individual pipes and would be best suited to models using large pipe groups or network-wide models. In an attempt to resolve the overdispersion problem, Zero-

inflated models were introduced. The ZIP GLM and ZIP negative binomial GLM were explored in the literature and benefited by treating the failures and non-failures separately in the model. However, ZIP models show mixed outcomes, suggesting they do not necessarily improve predictive accuracy. Grouping pipes is essential for the Poisson model to meet the appropriate assumptions. Careful consideration should be given to grouping schemes since they can affect the accuracy of the model and must consider the spatial scale for management decisions (Kleiner and Rajani, 2001). Alternatively, individual pipe models should make use of probabilistic approaches. Table 5-1 provides a summary of the main deterministic models discussed.

Table 5-1 Summary of the main deterministic pipe failure models.

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Shamir and Howard (1979)	Time-exponential model	-	-	-	Pipe-intrinsic	Network	Number of failures
Kettler and Goulter (1985)	Time-linear regression model	-	1959 to 1985	AC, CI	Pipe-intrinsic	Network	Failure rate
Kleiner and Rajani (2002)	Multi-variate Exponential model	-	1973 to 1998	CI, DI	Environmental Pipe-Intrinsic	Pipe Groups	Failure rate
Wang et al. (2009)	Five multiple regression models	432 km	1987 and 2001	CI, DI, HY, PVC	Pipe-intrinsic	Pipe Groups	Failure rate
Asnaashari et al. (2009)	Multiple regression & Poisson regression	554 pipes	10 years	AC, CI, DI, PE	Pipe-Intrinsic	Pipe Groups	Number of failures
Yamijala et al. (2009)	Exponential regression & Poisson GLM	<1,600 km (85,000 pipes)	2000 to 2005	AC, CI, DI, PVC, ST	Pipe-Intrinsic Environmental	Individual Pipes	Failure rate

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Kleiner and Rajani (2010)	Zero Inflated Poisson	146.6 km (1091 pipes)	1961 to 2006	AC, CI	Pipe-Intrinsic Environmental	Pipe Groups	Number of failures
Motiee and Ghasemnejad (2019)	Linear regression, Exponential regression & Poisson GLM	583 km	July 2004 to December 2007	AC, CI, DI, PE	Pipe-Intrinsic	Individual Pipes	Failure rate
Konstantinou and Stoianov (2020)	ZIP GLM and ZIP negative binomial GLM	374 km	2003 to 2016	AC, CI, DI	Pipe-Intrinsic Environmental Operational	Individual pipes / pipe lengths of 200 m	Failure rate

- = Unavailable information

1 GLM = Generalised Linear Model, ZIP = Zero-Inflated Poisson

2 CI = Cast Iron, DI = Ductile Iron, ST = Steel, AC = Asbestos Cement, PVC = Polyvinyl Chloride, PE = Polyethylene, HY = Hyprescon, RC = Reinforced Concrete

5.4.2 Probabilistic Models

5.4.2.1 Survival Analysis

The main disadvantages of deterministic models are the inability to accommodate left truncated data and their limited response, which led to the popularity of survival analysis. Survival analysis is capable of handling left-truncated data addressed analytically through the adaption of the likelihood function (see García et al. (2010) for further detail), and can be used to predict service life, the time between failures, or the development of consecutive failures over time (Wilson et al., 2017). This is achieved through predicting the elapsed time between an initiating event and a terminal event, or the time between events (requiring a particular condition, see Røstum (2000, p. 43)), revealing a pipes failure history (Alvisi and Franchini, 2010). The survival function and hazard function are fundamental aspects of survival analysis. Simply put, the survival function is the probability that a pipe survives longer than time, such that $S(t) = P(T > t)$, where $S(t)$ is the pipe at time t , and T represents the time until the occurrence of an event. The hazard function $h_i(t)$ is the probability of a pipe i failing within that time, formally expressed as:

$$h_i(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad \text{Eq.5-3}$$

Where $f(t)$ is the probability density of function T and $F(t)$ is a cumulative distribution function (Rodriguez, 2007). The hazard function is represented as either a constant, increasing decreasing or bathtub shaped (Røstum, 2000), where a hazard of greater than one signifies an increased hazard and a decreased life expectancy of the pipe. The survival rate and failure of a pipe are illustrated in Figure 5-2 taking an exponential shape.

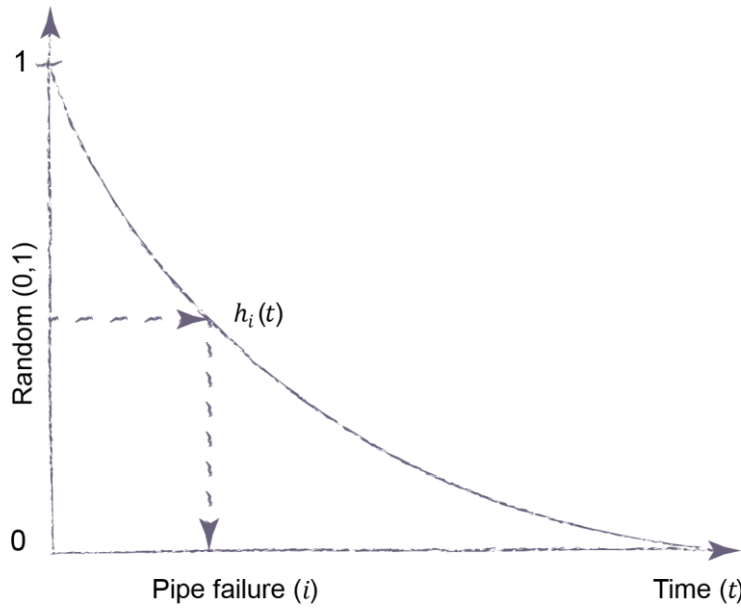


Figure 5-2 Illustrative example of a survival curve (adapted from Le Gat and Eisenbeis (2000)).

The proportional hazard (PH) is a standard model introduced by Cox (1972) and focuses directly on the hazard function. Early pipe failure models using the Cox PH were developed by Andreou et al. (1987a, 1987b), including a semi-parametric model developed to determine slow and fast failure phases. The proportional hazard and survival function are fundamental elements that define the Cox PH, which takes the general form:

$$h_i(t/X) = h_0(t) \times \exp(\beta_1 x_1 + \dots + \beta_i x_i), \quad \text{Eq.5-4}$$

with the assumption that the baseline hazard h_0 varies over time t , equivalent to the total hazard rate when the covariates equal zero (Kabir et al., 2016; Wilson et al., 2017). The model predicted the probability of failure for individual pipes applied to two water utility networks in the US. The authors reported that the failure rate increased with every subsequent failure up to the third failure, after that remaining constant. This was one of the first studies understanding the relationship between failures over time. The authors made other practical

observations; smaller diameter pipes fail more often; pipe pressure has a low correlation with pipe failures and pipes that fail early in their life cycle outperform those that fail later. The main limitation with the study was the use of varying pipe lengths, with some pipes over 1km, a length where conditions affecting pipe failure can vary greatly. A further limitation was the use of a single pipe material, which does not represent the whole network, the lengths of pipe vary greatly in the study, which can affect the break rate. The main advantage of the Cox PH is that the hazard function assumes no form. Therefore, it is semi-parametric, which makes evaluating the effect of covariates on the hazard easier. The authors developed their model specifically for this reason.

When the objective is to predict future failures over a time horizon and of a particular form, a parametric model is considered more appropriate, such as the Weibull PH (Røstum, 2000). The Weibull PH has also been widely used for pipe failure models (Alvisi and Franchini, 2010; Mailhot et al., 2000; Røstum, 2000) and is a simpler parametric model that assumes the natural log of time to failure is linearly related to the covariates (Le Gat and Eisenbeis, 2000). Le Gat and Eisenbeis (2000) used a Weibull PH to understand the interarrival time of pipe failures, using short maintenance records to forecast failures for different pipe materials. In this study, the authors used a Monte Carlo simulation to generate random data in the model, which helped to improve the predictions. Applied to two datasets in Charente-Maritime and Lausanne, the model worked well in the first but failed to predict the second dataset well due to inherent increased pipe degradation, inadequate data, and missing environmental variables. Mailhot et al. (2000) used a failure order model applied to a 21-year (1976-1996) data set in Chicoutimi near Quebec City, with a WDN of approximately 352 km (2096 pipes). The authors used four combinations of the Weibull and exponential distributions, comparing different installation periods on the shape of the hazard function, specifically considering short failure records and left truncation of the data. The results showed that the installation period strongly influenced pipe failures and concluded that an explicit unrecorded period of data must be included

when failure history is absent. However, the model was limited by considering subsequent failures to follow a single variable exponential type distribution. Alvisi and Franchini (2010) applied both models from Le Gat and Eisenbeis (2000) and Mailhot et al. (2000), applied to a WDN in Ferrara, Italy, with a failure record of seven years and ~ 3500 failures. Comparing groups of pipes, the authors found that the models performed similarly in predicting failures, yet the model by Mailhot et al. (2000) was restrictive due to only considering pipe installation period, where large networks have more data availability. Scheidegger et al. (2013) used a Weibull distribution to model the first failure and all subsequent failures using an exponential distribution to facilitate smaller datasets by simplifying the approach to accommodate fewer variables.

Debón et al. (2010) compared different models to understand which performed more reliably when predicting the time of failure. Three models were chosen, including a Cox PH model, Weibull accelerated lifetime model and a GLM. The results led the author to conclude the minimal difference between the Cox and Weibull models, but comparing the Cox model to the GLM, the GLM showed a better fit when comparing the Receiver Operating Curve (ROC) (0.77 and 0.83, respectively), a measure of accuracy that would later be used in many more studies. Unusually, the authors noted that pipes installed under sidewalks were less prone to failure. Comparing the different survival analysis models and a Poisson model using a WDN consisting of 31,662 individual pipes (4,281 km) grouped by similar characteristics, Kimutai et al. (2015) reported a lower RMSE for the Weibull PH for two out of three materials suggesting it performed better than the other models on this dataset. The Weibull model managed to capture the failure phases of metallic pipes more appropriately. Comparing the Weibull to a machine learning gradient boosting model, Snider and McBean (2019) found that the machine learning outperformed a Weibull PH model on a WDN with a long failure history (1960 – 2006). The dataset had a failure period of 55 years, with 47 % of the pipes having at least one failure record, which was reported as

advantageous for the models. The machine learning predicted failures early because it was unable to account for censored data like the Weibull PH model.

5.4.2.2 Logistic Regression

Logistic regression calculates the probability P of failure over the joint probability distribution $P(x, y)$ through the logit function (Louppe, 2014), where a probability of one indicates a failure and a probability of zero no failure, and all probabilities must be within this [zero-one] interval (St. Clair and Sinha, 2012). As previously discussed (see Section 4.1.3), probabilistic models such as logistic regression have proved to be more useful when compared to Poisson, linear and exponential models, and returning the probability of a failure is enough to inform management decisions (Yamijala et al., 2009; Motiee and Ghasemnejad, 2019), yet these studies are limited in their linear assumptions. Further studies used the logistic function in the Generalised Additive Model (GAM), rarely used or discussed in pipe failure literature, but valuable since it extends the GLM by including a smoothing function f_i that can measure arbitrarily non-parametric relationships, such that a simple form might be $y_i = \beta_0 + \beta_i x_i + f_i(x_i)$ (Barton et al., 2020). The semi-parametric models are advantageous over parametric models since they are adaptable to the complexity of pipe failure data. Chen et al. (2019) performed a comparison of models at two levels, the road level and the census level (2.3 km² grid areas), predicting the probability of failure using data recorded between 2010-2014. The study was one of the first to include a GAM with logistic regression function and found that it performed best compared to several other deterministic and machine learning models at the road level, being able to identify areas of a high probability of failure more accurately. The authors acknowledged the limited accuracy overall due to a lack of asset-level information, the short prediction interval, the limited network information, and the limited failure records. Data imbalance was addressed, but the full details were not provided. Variable selection was performed for the models but found to have a limited effect on the overall accuracy. Compared to historical failure rate-based ranking models, the

GAM performed better and was preferred since it is easier to develop and computationally efficient when compared to machine learning models.

5.4.2.3 Bayesian Models and Expert Knowledge

Due to a lack of suitable data, researchers started to explore Bayesian models for their ability to incorporate prior beliefs or expert knowledge by constructing a probability distribution that describes the model uncertainty prior (prior distribution) to the data from the experiment to provide a formal estimate of failure rate (posterior distribution), such that $Posterior_i = \frac{Prior_i \times Likelihood_i}{\sum_i Prior_i \times Likelihood_i}$ (Kabir et al., 2016; Watson et al., 2004); a significant advantage given the complexity of WDNs and their local geographic nature (Deadman, 2010). Watson et al. (2004) were one of the first to employ a hierarchical Bayesian model to simulate an object-orientated event, assuming the underlying failure rates for all pipes are drawn from the same prior distribution, called the hyperprior. Adjusting the hyperprior variance determines the individual failure information, where a small variance implies pipes have similar failure rates and a high variance that pipes are different. The model was tested on two pipes over 50 years, updating the model every five years. Compared to a natural estimation (Poisson distribution), the Bayesian model with gamma prior distribution provided a better failure rate estimate. More recent developments combine survival analysis with the Bayesian model. Kabir et al. (2016) developed a Bayesian Weibull PH Model (BWPHM) using Bayesian model averaging for parameter estimation. The covariates included length, diameter, vintage, land use, freezing index, rain deficit, soil resistivity, and corrosivity. When comparing the BWPHM against a Cox PH and Poisson model for a medium-sized WDN (4,281 km) on individual pipes and data collected between 1956-2013, the results were favourable; Root Mean Squared Error (RMSE) of 8.65, 11.73, and 13.08 respectively for cast iron pipes. However, the Bayesian averaging benefits are not clear since the Weibull approach has been shown to outperform the Poisson and Cox PH models in previous studies (Kimutai et al., 2015). The model showed soil resistivity and soil corrosion as the most important variables.

Further developments included the use of Bayesian Networks (BN), capable of capturing conditionally dependent and independent relationships through a directed acyclic graph structure with edges and parent and child nodes and links representing the causal dependencies between them, therefore explicitly quantifying uncertainties. A full description of the model is aptly provided by Tang et al. (2019). This complex approach can explicitly quantify uncertainties, identify complex relationships, and use engineering knowledge and data. In a study by Francis et al. (2014), a BN was applied to a WDN in the Mid-Atlantic using a short data set collected between 2010 and 2011. The data included 3,686 failure records, which were poorly located on the street level. To overcome this, the pipes were grouped at the census level. The results showed sustained challenges for the model, which the authors suggested were due to the monthly prediction interval and zero-inflation problem, and the inability of the BN model to discriminate between signal and noise due to limited information in the distributions. The study was limited by the unavailability of important pipe data, and therefore census, demographic, weather, and soil variables were used and standardised to deal with scaling issues; however, not many studies have tried to incorporate such population characteristics. The authors noted that many BN structures assume all random variables to be discrete, and so variables were discretised, which can have a significant impact on model accuracy, and in this study may have biased the results because of the zero-inflation problem. The authors also suggested that no single technique is better for obtaining the joint distribution from expert knowledge. Tang et al. (2019) applied an automated BN based on historical data and a guided BN utilising historical data, literature, and expert knowledge. The methods were applied to individual pipes collected across many WDNs across the UK. The authors reported poor accuracy (Area Under Curve (AUC) of 0.7860 and 0.702, respectively) because of poor data quality, which influenced the Bayesian approach despite expert knowledge. Although not reported in this study, expert knowledge can be problematic with issues arising from different expert knowledge providing different posterior distributions (Kabir et al., 2018). Scholten et al. (2013) suggested that experts cannot be expected

to know the distributional form or the correlation of variables with failure, and so developed a means to elicit expert knowledge into a form of stated quantiles, using an approach that minimised cognitive bias. The method used several experts to define the likely probability or frequency of relationships. Using the Bayesian inference and Weibull model to predict service life, the results showed improved service life estimation under scarce data. The authors compared the Weibull distribution to lognormal and gamma and found the Weibull more effectively approximated the non-parametric curve. However, these distributions take standard recognisable forms, which is rarely the case, therefore some studies have used the Monte Carlo Markov Chain (MCMC) technique to solve this problem (Economou et al., 2008; Lin and Yuan, 2019).

5.4.2.4 Non-Homogenous Poisson Process

The Non-Homogenous Poisson Process (NHPP) is a probabilistic modelling approach that captures pipe deterioration, expressed as a failure rate using a monotonic function of time such as the log-linear or log-power relationship. A critical feature of the model is that it relaxes the Poisson process and can, therefore, allow the failure rate to vary over time, capturing the deterioration mechanism (Economou et al., 2012). Røstum (2000) applied the NHPP to a WDN of 808 km in Trondheim, Norway, favouring the NHPP compared to a Weibull PH that overestimated the number of predictions. Economou et al. (2008) developed an NHPP and a Zero-Inflated NHPP (ZINHPP) using Bayesian parameter estimation and random effects. The NHPP process was flexible and captured the deterioration of pipes well, yet the ZINHPP model did not outperform the NHPP since the data had adequate information on pipe failures. Economou et al. (2012) further updated their initial study to account for the over-parametrisation. Here the model was applied to a case study of two medium sized WDNs (532 pipes and 1,349 pipes, respectively) with different observation periods of 11 years (1990-2001) and 41 years (1962-2003). The authors found that the ZINHPP model performed better than the NHPP when comparing the Deviance Information Criteria (DIC) for one dataset (New Zealand data DIC 344 and 515,

respectively), allocating more probability to no failure. Both models showed similar results for the second dataset, potentially because there were more failures (only 22 % non-failures compared to the first dataset of 85 % non-failures) in the data; therefore, the data was less imbalanced. Neither Røstum (2000) nor Economou et al. (2012) used environmental variables limiting the study, and the main limitation of the NHPP is that the influence of previous failures cannot be modelled.

Kleiner and Rajani (2010) applied a ZIP model for a Non-Homogenous Poisson Process (NHPP) (see Section 5.4.2.4) to a WDN in western Canada, comprising 1091 pipes (length of 146.6 km) with a failure history between the dates 1961 and 2006. The variables used included material, diameter, installation year, length and soil and weather. The model successfully predicted the number of failures in each group for the five-year training data set ($R^2 = 0.61$) but failed to accurately predict failures for individual pipes ($R^2 = 0.43$), overestimating the number of failures, which may have arisen due to many pipes having zero failure, and a few pipes having many failures. The study was performed on only 150 mm diameter unlined cast iron pipes, which limits the study, and whilst weather variables were used in training, they cannot be used to forecast pipe failures (Kleiner et al., 2010). Further developments for time-dependent weather variables (temperature, water temperature and freezing index) was undertaken by Rajani et al. (2012) to understand the influence on pipe failures and the best time step for predictive accuracy. Predicting the mean failure rate for three data sets in the USA and Canada, the findings showed all influenced failure rates, but the temperature was vital since it is usually available. Comparing a prediction interval of 5, 15, 30, 60 and 90 days, the results showed improved accuracy with more extended periods, ultimately deciding on 30 days to be the most appropriate. However, the study modelled six large groups of pipes and did not address individual mains within the group, and the authors noted poorer results with shorter intervals. Kleiner and Rajani (2012), in a continuation of their research (Kleiner and Rajani 2010), used an NHPP model to predict annual failures from a long failure history (1961-2006)

and a medium WDN of 146.6 km. Comparing four alternative models (an ordered list heuristic model, naïve Bayes, logistic regression and NHPP) and using important weather variables, the authors noted that the large data set introduced noise which led to difficulty in obtaining meaningful results, and weather variables showed limited influence due to the pipe depth (2.4 m). The authors also noted time lags between the occurrence of failures and their discovery, potentially having discernible impacts on short prediction intervals. Amongst the models, only the NHPP could predict the number of failures. Therefore, the models were compared by their ability to rank pipes expected to experience the highest number of failures. The results revealed that no model was superior in ranking pipes, but the NHPP was better at considering the time-dependent covariates and was advantageous since it can predict the number of failures.

5.4.2.5 Comments

Probabilistic models are flexible, can handle randomness, capture the complexity of multiple variables, and are capable of various response types. Probabilistic models are appropriate for modelling individual pipes, providing helpful information for rehabilitation and replacement. Survival analysis and, in particular, the PH model is insensitive to left-truncated data and can measure different failure phases of a pipes life cycle and times between failures. In this respect, survival analysis is unique since it can determine the best time to replace a pipe over a long prediction interval, considering the likely number of failures and the financial comparison between replacement or repair. Much of the literature has reported that the Weibull distribution suitably describes time-to-first failure, and the exponential distribution suitably describes subsequent failures. The main disadvantages of survival analysis include the complexity and difficulty handling short failure recording periods, where explicit unrecorded periods of data should be included (Mailhot et al., 2000).

Logistic models offer a simple mathematical framework that water companies can readily use and provide a probability of failure between 0 and 1 or a classification approach of failure or no failure, helpful for management decisions. Yamijala et

al. (2009) and Motiee and Ghasemnejad (2019) have shown that logistic regression can work with short failure records, and is favourable compared to the linear, exponential and Poisson approach when modelling individual pipes. However, all studies show limitations in data quality and focus on short term intervals and short failure records, which may have resulted in the poor performance of other models. The main disadvantage is the assumption that the variables are related to pipe failures in a linear nature, which fails to describe complex non-parametric relationships. Therefore, Chen et al. (2019) explored the logisitic approach using a GAM, which almost compared in performance to machine learning. However, other probabilistic models are more likely to outperform logistic regression, such as Bayesian models.

Bayesian models offer a dynamic approach to modelling failures, uniquely using a graph structure that can describe complex non-parametric relationships, are easy to interpret and can explicitly quantify uncertainty. Bayesian models are important given they can also include expert knowledge, advantageous given the lack of information in many datasets, but the joint distribution must be carefully chosen, and there is no preferred technique over others for obtaining this (Francis et al., 2014). Collecting expert opinions can be onerous (an interesting approach has been used by Scholten et al., (2013) and introduces bias, since it is difficult for managers to establish the most important effects on pipe failures when they are often responsible for small areas of the entire network, instead of having a holistic view (Deadman, 2010). Care must also be taken with insufficient failure data, since estimated parameters will be more sensitive to these prior information (Dridi et al., 2005). Bayesian models have been combined with survival analysis or Non-Homogenous Poisson models as a means of parameter estimation, which has suggested improved accuracy. Like survival analysis, Bayesian models are computationally complex, which dissuades their use somewhat. The NHPP model is advantageous since it is flexible and allows failure rates to vary over time, capturing the deterioration mechanism in water pipes, which is useful when the actual time of a pipe failure is unavailable. Care must be taken when

discretising data since this may cause bias in zero-inflated datasets (Francis et al., 2014). However, the NHPP can account for the zero-inflation problem, yet little evidence suggests that zero-inflated models more accurately represent pipe failures. Table 5-2 shows a summary of the main probabilistic models discussed.

Table 5-2 Summary of the main probabilistic models discussed.

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Andreou et al. (1987a, 1987b)	Cox PH	-	-	-	Pipe-intrinsic Operational		Probability of failure
Le Gat and Eisenbeis (2000)	Weibull proportional hazard model	1243 km (1212 pipes)	Nine years	AC, CI, PVC, ST	Pipe-intrinsic Environmental	Individual pipes	Number of failures
		688.7 km (6966 pipes)	Since 1926		Pipe-intrinsic Operational		
Mailhot et al. (2000)	Weibull exponential model	352 km (2096 pipes)	1976 to 1996	-	Pipe-intrinsic Environmental	Pipe groups	Failure age probability
Røstum (2000)	NHPP Weibull PH	7,627 pipes	1988 to 1996	CI, DI, plastic, others	Pipe-intrinsic Environmental	Individual pipes	Number of failures
Watson et al. (2004)	NHPP	-	-	-	Pipe-intrinsic Expert Knowledge	Individual pipes	Time-to-failure

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Economou et al. (2008)	NHPP/ ZINHPP (Bayesian framework)	1349 pipes	1969 to 2003	CI	Pipe-Intrinsic	Individual pipes	Failure rate
Yamijala et al. (2009)	Logistic regression	<1,600 km (85,000 pipes)	2000 to 2005	AC, CI, DI, PVC, ST	Pipe-Intrinsic Environmental	Individual pipes	Probability of failure
Debón et al. (2010)	Cox Ph Weibull accelerated lifetime model GLM	32,387 pipes	2000 to 2006	CI, DI, PE	Pipe-intrinsic Environmental Operational	Individual Pipes	Time-to-failure
Alvisi and Franchini (2010)	Weibull proportional hazard model Weibull exponential model	2,400 km (23,000 pipes)	2000 to 2006	AC, CI, PE, PVC, ST	Pipe-intrinsic Environmental	Pipe groups	Time-to-failure (inter-arrival time)
Kleiner and Rajani (2010)	Zero-Inflated Non-homogenous Poisson Process	146.6 km (1091 pipes)	1961 to 2006	AC, CI	Pipe-Intrinsic Environmental	Pipe groups	Probability of failure

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Economou et al. (2012)	ZINHPP	1,349 pipes 532 pipes	1962 to 2003	CI	Pipe-intrinsic Operational	Individual pipes	Failure rate
Rajani et al. (2012)	NHPP	2200 km (16,383 pipes)	1972 to 2001	CI	Environmental	Pipe groups	Number of failures
Kleiner and Rajani (2012)	Heuristic model, Naïve Bayes, Logistic regression and NHPP	370 km	1962 to 2006	AC, CI, DI	Pipe-Intrinsic Environmental	Individual pipes (within a group)	Rank (Pipe expected to experience the highest number of failures)
Scholten et al. (2013)	Weibull, lognormal and gamma models	322 km (3,643 pipes)	2001 to 2010	AC, CI, DI, PE, ST	Pipe-intrinsic Environmental Expert knowledge	Pipe groups	Time-to-failure (inter-arrival time)
Kimutai et al. (2015)	Weibull PH model Cox PH model Poisson model	4,281 km (31,662 pipes)	1956 to 2013	CI, DI, PVC	Pipe-intrinsic Environmental	Pipe groups	Time-to-failure (inter-arrival time)

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Kabir et al. (2016)	Bayesian Weibull PH Cox-PH Poisson model	4,281 km (49,531 pipes)	From 1956	CI, DI, PVC, other	Pipe-intrinsic Environmental	Individual pipes	Time-to-failure (inter-arrival time)
Snider and McBean (2019)	Weibull PH xgboost	30,000 pipes	1960 to 2005	AC, CI, DI, PVC	Pipe-intrinsic	Individual pipes	Time-to-failure (inter-arrival time)
Motiee and Ghasemnejad (2019)	Logistic GLM	583 km	2004 to 2007	AC, CI, DI, PE	Pipe-Intrinsic	Individual pipes	Probability of failure
Tang et al. (2019)	Bayesian network Guided/ learning Bayesian network	-	1980 to 2017	AC, CI, DI, GRP, PE, PVC, ST	Pipe-Intrinsic Environmental Expert Knowledge	Individual pipes	Probability of failure
Francis et al. (2014)	Bayesian Belief Networks	-	2010 to 2011	-	Pipe-intrinsic Environmental Census	Pipe groups (census)	Probability of failure

- = Unavailable information

1 GLM = Generalised Linear Model, NHPP = Non-Homogenous Poisson Process, PH – Proportional Hazard, ZINHPP = Zero-Inflated Non-Homogenous Poisson Process

2 AC = Asbestos Cement, CI = Cast Iron, DI = Ductile Iron, GRP = Glass Reinforced Plastic, PE = Polyethylene, PVC = Polyvinyl Chloride, Concrete, ST = Steel

5.4.3 Machine Learning

Recent developments in machine learning have greatly expanded the capabilities of statistical models and are now more commonly used than traditional models due to their improved predictive accuracy (Giraldo-González and Rodríguez, 2020). The advantages of machine learning include removing unnecessary mathematical processing steps in some instances (decision trees (Winkler et al., 2018)), variable selection through shrinkage estimators, and tuning of interaction terms through cross-validation, which offers greater flexibility than simple regression models. For structured data, such as that seen in pipe failure data, supervised grey-box models are more appropriate (see Figure 5-3). Supervised models can be tuned for improved model accuracy and are interpretable through variable importance measures and partial plots (Wols et al., 2019), which is more appealing than black-box approaches to industry professionals (Barton et al., 2021). There are various studies in the literature that have applied machine learning and explored its effectiveness. Commonly used models in the literature include Artificial Neural Networks (ANN), Support Vector Machines (SVM), Evolutionary Polynomial Regression (EPR) and more recently, tree models.

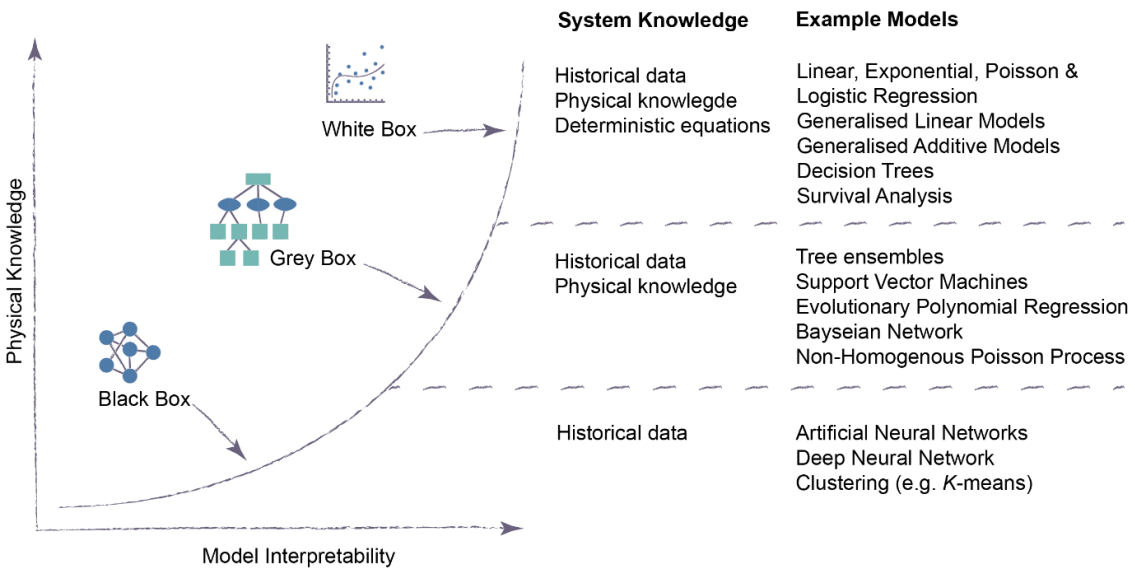


Figure 5-3 Pictorial representation of white-box, grey-box, and black-box models, showing how the greater physical knowledge exists, the better the interpretability of pipe failures (adapted from Rezania (2008)).

5.4.3.1 Clustering

Pipe failures can be modelled based on spatiotemporal relationships, assuming that more failures presented within proximity to one another indicate distress in the network. Spatiotemporal relationships in failures were first noted by Clark et al. (1982) and Goulter and Kazemi (1988), reporting that consecutive failures were related to the first. These early studies provided essential insights not seen before. However, the studies did not constrain the relationships along with a pipe network, which has distorted the results. De Oliveira et al. (2011a) expanded this further by suggesting that spatiotemporal relationships should be network restricted and used Dijkstra's algorithm (Dijkstra, 1959) to ensure that the constraints of the pipe network have been considered. In the study, the authors used Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to find areas of the network underperforming by identifying hot spots with high failure rates. The authors conclude that poor repairs of failures might be responsible for producing new failures. Chen and Guikema (2020) used locally weighted DBSCAN and two Poisson based models to determine the historical cluster of

breaks and used the results as an exploratory variable in pipe failures. The results suggested that DBSCAN results in higher precision clusters and was then used in different regression models, including GLM, GAM, Random Forest (RF), and Gradient Boosting Trees (GBT), with each model including and excluding the cluster variable. The results revealed improved AUC scores for all models, including the cluster variable by 6.2, 3.1, 3.3 and 4.6 %, respectively. The authors note that it is important to normalise data (to achieve a common scale of data) when performing cluster analysis and reported that overall accuracy measures do not always reflect the success of a model. Instead, the success should be based on the ability to prioritise pipes at high risk and therefore use a rank-based performance measure. Aslani et al. (2021) continue the use of failure clusters by using Getis-Ord (Getis and Ord, 1992) to recognise spatial clusters of pipe failures in the city of Tampa, Florida, for a five-year dataset (2015-2020). The results of the cluster analysis were used as a variable for various machine learning models. The data were prepared using multiple imputations for missing values, and categorical data were processed into numeric data using one-hot-encoding. The accuracy of the models was presented as RMSE and rank-ordered break (similar to Chen and Guikema (2020)) and suggested poorly fit models, which is likely to be a result of the spatiotemporal extent of the model leading to data imbalance. However, the authors found the cluster variable to be significant and concluded that a boosted regression tree was the most reliable model.

K-means clustering assigns data samples into clusters around a centroid, using a function that iteratively minimises the dissimilarity between data (the Euclidean distance). The process then recalculates and moves the centroid reassigning the data, repeating the process until an optimal balance is achieved (Kleiner and Rajani, 2012) (see Figure 5-4). *K*-means clustering is typically applied during pre-processing failure models to group pipes with precision (Farmani et al., 2017), modelling each group individually to improve failure prediction accuracy. There is no specific study that suggests *K*-means clustering outperforms other grouping methods. However, Kakoudakis et al. (2017) compared an EPR model to a *K*-

means clustered EPR model, using a UK city WDN. The results showed better predictive accuracy in the clustered models when categorising pipe failure rates; 55 and 85 % accuracy, respectively, for low, medium, high, and very high-risk categories. Cluster analysis is a valuable tool for identifying why failures may occur, yet it cannot predict the future pipe state since there is no response variable; therefore, it is not strictly in itself a failure model.

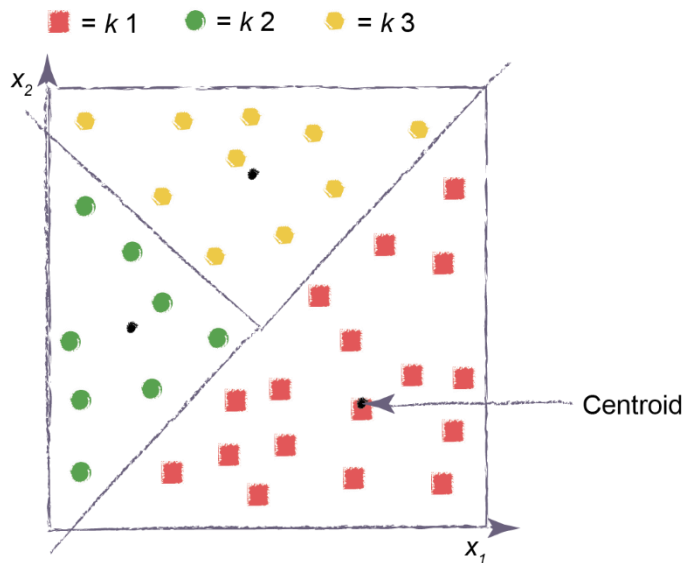


Figure 5-4 Illustrative example of *K*-means clustering in a two-dimensional space, depicting three clusters of the n data samples, based on Euclidean distance to the cluster centroid.

5.4.3.2 Artificial Neural Networks

ANN was one of the first machine learning models adopted in pipe failure models. It was used to improve the sometimes poor predictive accuracy seen from traditional models and is often likened to a biological neural network based on how the algorithm processes information (Kalogirou, 2004). Considering this, ANNs comprise several process units known as neurons l arrayed in each layer. Each neuron within a layer is connected to each neuron within the joining layer. The neuron calculates a weighted sum w_{ij} of each input variable from the first layer to the second layer (hidden layer), which is made of activation units that

perform the activation function f , transforming the data into non-linear outputs (such is the hyperbolic symmetric sigmoid function (Kutyłowska, 2015), yet many other functions exist). The third layer is a single activation unit that takes weighted outputs from the second layer and predicts y . ANN takes the general form:

$$y_j = f\left(\sum_{i=1}^l w_{ij}x_i + a\right) \quad \text{Eq.5-5}$$

Where a is a threshold (Wei et al., 2018). The process is illustrated in Figure 5-5.

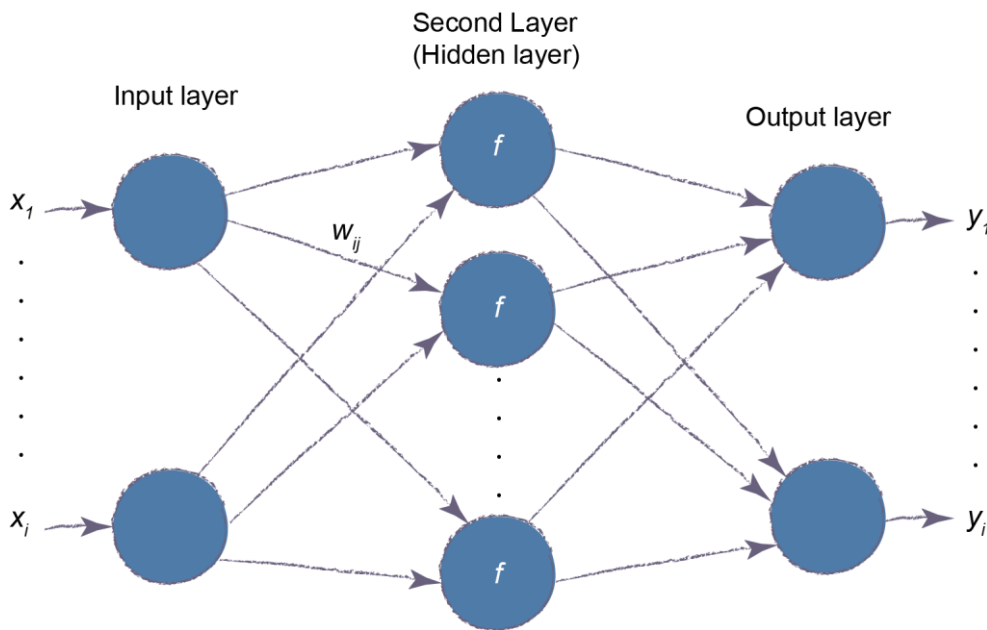


Figure 5-5 Simple Artificial Neural Network architecture illustrating only three network layers (adapted from (Wei et al., 2018)).

Some of the early ANN models were combined with fuzzy logic; an algorithm used to handle imprecise information and approximate reasoning. In this respect, fuzzy logic is a technique that can return a logic gradient of between 0 and 1, and when combined with ANN, provides a human-style logic with the neural style learning structure (Kleiner et al., 2004). Fuzzy logic alone has been used to determine a pipes deterioration procedure (Kleiner et al., 2004) and translate inspection investigations into pipe condition ratings (Rajani et al., 2006). One of the first ANN models was developed by Christodoulou et al., (2003), who used

an ANN in a hybrid model with fuzzy logic. The model was used with Weibull distributions and Kaplan-Meier survival curves to explain the time-to-failure on groups of pipes. Twelve variables were used in the model and applied to a network in New York covering a total length of 365 km. Pattern recognition using fuzzy logic was employed to rank the risk of failure and finally assembled in a GIS platform for visualization. The authors suggested that the model successfully ranked the risk of failure and identified material, diameter, length, and the number of failures as the most influential variables. Building on this model, Christodoulou and Deligianni (2010) further developed the use of ANN model again with fuzzy logic for two small WDNs (365 km in New York, and 795 km in Limassol, Cyprus) of individual pipes using only five significant variables identified as important from previous studies, including traffic (not typically used in models). The fuzzy logic approach is linked to a management tool that predicts the risk of failure over time, a useful output for industry. The authors attest to the success of the model, which was rolled out into a pilot scheme in two cities and highlighted the strength of ANN to facilitate incomplete time-sensitive multi-parameter data suitably. Tabesh et al. (2009) used an Adaptive Neuro-Fuzzy Inference System (ANFIS), a method that obtains the if-then rule automatically through the learning capability of the ANN model instead of through fuzzy rules. The ANFIS was compared to an ANN and applied to a network of 580 km of pipe and five pipe-intrinsic variables. The authors concluded that the ANN model provided more reliable results since it had a lower magnitude of variation in the results.

Other notable studies include Jafar et al. (2010), who employed different ANN models (split by material and a global model for comparison) to predict failure rates and replacement times of individual pipes on a dataset of 162 km (4,862 pipes) in northern France between 1991 and 2004. The authors used 11 variables that were normalised (since ANN models can be sensitive to outliers) to their minimum and maximum values after categorical data were converted into separate variables using dummy coding, including soil and pressure. A variety of variables were used, with length, age, and the number of previous failures the

most significant. Like other studies, the global model performed well; for cement, plastic, metallic and global models, R^2 of 0.589, 0.671, 0.522, 0.671 for the testing data. The main advantage of this research was using a benefit index for optimising investment, identifying 5 % of pipes in the global model for replacement, potentially reducing failures by 53 %. Nishiyama and Fillion (2014) predicted the total number of failures over a two year and five-year period applied to a WDN in Ontario, Canada. Pipe failure records were collected between 1998 and 2011 and included diameter, age, length, and soil type variables. The results showed an approximate AUC of 0.78 and overall accuracy of 85 %, yet the true positive predicted only ~ 40 %. Harvey et al. (2014) only found acceptable results when predicting time-to-failure for asbestos cement, cast iron and ductile iron, with an adjusted R -value of 0.70, 0.70 and 0.81, respectively, when using data from a medium WDN of 5,850 km over a pipe failure dataset recorded between 1962 and 2005.

Kakoudakis et al. (2018) advanced the use of ANN by using the model to distinguish between a day with and without failure, a short prediction interval not previously attempted. The model used six clusters of pipes determined using k-means clustering, cross-validation for tuning and a binary response. The authors observed an AUC measure of 0.184, and the results help alert water companies to potential failure areas. Sattar et al. (2019) proposed a novel Extreme Learning Machine model (ELM), an alternative to ANN with a basic structure of a single layer network, making it computationally fast and has a high generalisation capacity. The authors predicted time-to-failure using pipe coatings, material, length, and diameter and applied the model to a network in Scarborough, Greater Toronto, which contains 6,342 water mains over 1000 km, and failure records were collected from 1962. Comparing the model to feed-forward ANN, SVM and non-linear regression, the ELM had superior predictive accuracy. One significant advantage of this study was the long failure records, which resulted in 44 % of the pipes having at least one failure.

When comparing ANN to other models, there are, however, mixed results. Asnaashari et al. (2013) predicted the failure of two models and found ANN (R^2 of 0.94) to have a better predictive capability than multiple linear regression (R^2 of 0.75, or 0.63 with cross-validation) when applied to a 5,850 km WDN with a long failure record history (since 1960). Giraldo-González and Rodríguez (2020) compared the probability of failure in four machine learning models and found that neural networks poorly predicted the minority class (failures). Over-fitting is the main disadvantage of ANN since the network must be identified *a priori* (e.g. the number of neurons, activation functions, training epochs, hidden layers, learning rate and momentum term) (St. Clair and Sinha, 2012). In a review of literature, Wilson et al. (2017) excluded ANN models, suggesting they had high data requirements and was therefore unsuitable for large diameter pipe models.

5.4.3.3 Evolutionary Polynomial Regression

EPR is a data-driven hybrid regression technique that uses mathematical structures based on evolutionary computing developed by Giustolisi and Savic (2006) and is widely used for pipe failure modelling. EPR is a two-step process; firstly, it looks for the best model structure using a Multi-Objective Genetic Algorithm (MOGA); secondly, it finds the parameter estimation using the ordinary least squares method (Berardi et al., 2008). In this respect, EPR builds symbolic models by integrating features of genetic programming (Xu et al., 2011), numerical regression, and symbolic regression, establishing parsimony between the polynomial terms and the variable. The general EPR form is:

$$y = \sum_{j=1}^m F(X, f(X), c_j) + b \quad \text{Eq.5-6}$$

Where c_j is the constant, f is a user defined function and F an EPR constructed function (Giraldo-González and Rodríguez, 2020).

Berardi et al. (2008) used an EPR model to predict pipe failures and derive performance indicators as a means of representing the propensity of a pipe to

fail. The study assessed individual pipe criticality, defined as the failure likelihood and the expected damage. Pipes were grouped by diameter and age and only included pipes < 250 mm in diameter since the model was only economical for small diameter pipes. The variables used in the model included age, length, diameter, and the number of properties. The accuracy measure for EPR was the Coefficient of Determination (CoD). The results showed that the EPR model overall fit with a CoD of 0.822 but over-estimated failure rates in pipes with no failure history because of the limited failure history. The approach was simple and produced understandable relationships between failures and variables, yet the study did not validate the predictions. The main disadvantage in this study was that the timing of failures was unknown. Laucelli et al. (2014) used EPR to understand dynamic time-related climate variables. Using Rajani et al. (2012) as a starting point, the authors extended the range of dynamic temperature variables and applied the EPR on a data set for Scarborough (Ontario, Canada) focusing on 150 mm diameter CI pipes with a failure period between 1962 and 1985. Overall, the model showed reasonable accuracy (CoD = 0.79); however, when investigating the relationship between climate data, the authors found good accuracy for failure rates during the winter, but not for the summer, since the phenomena that result in high failure rates for cast iron pipes during the winter cannot explain the failures during the summer. However, the models determined warm and cold seasons by applying a threshold for the number of freezing days rather than distinct seasons, which may have influenced the results. The sensitivity of the model to these thresholds was not explored in the study.

Farmani et al. (2017) and Kakoudakis et al. (2017) explored the use of *k*-means clustering with EPR models, and both studies identified an improved performance when grouping the pipes and modelling each group with a separate EPR model. The authors presented the response as failure rates, and the risk of failure was categorised successfully by using Natural Jenks, an approach rarely used for pipe failure models. This premise was continued in a further study by Kakoudakis et al. (2018), where the authors used EPR with *k*-means clustering to partition the

training data into different groups for the EPR. The pipes were divided into six clusters based on diameter and age, showing reasonable results from very low to very high-risk categories for long term predictions (with an accuracy of between 46 and 87 %) and good performance for the short-term probability model (AUC of 0.814). Comparing EPR to other models, Giraldo-González and Rodríguez (2020) found that Poisson regression outperformed EPR (R^2 of 0.927 and 0.885, respectively) when predicting groups of pipes for a medium-sized WDN, again the authors used k-means clustering to create pipe groups and observed its benefits. EPR models are intuitive and can include expert judgement in the process. The approach has mainly been used for groups of pipes and not for individual pipe models (Berardi et al., 2008).

5.4.3.4 Support vector machines

SVMs are linear models that maximise the width of separation between data classes while minimising the error (Louppe, 2014, p. 22). Good separation between the data classes is achieved when a large distance, also known as the margin, exists on either side of a hyperplane (see Figure 5-6). A linear hyperplane satisfies $w^T x - b = 0$, and the upper and lower margins $w^T x_i - b \geq 1$ if $y_i = 1$ and $w^T x_i - b \leq -1$ if $y_i = -1$ respectively, where w is the weight of the variable space, and b is a bias term. SVMs can implicitly model non-parametric data using a kernel function (standard kernel functions commonly used include polynomial, hyperbolic tangent, and radial), mapping the variables into a high dimensional space, separated by multiple hyperplanes (Louppe, 2014; Robles-Velasco et al., 2020). SVM is useful for handling outliers by allowing misclassifications, letting some data fall within the margin to obtain better results, but limiting their use in the model via a Slack variable ϵ_i applied to the outlier (Giraldo-González and Rodríguez, 2020). Put together for a range of input variables, and the basic SVM model is (Robles-Velasco et al., 2020):

$$y_i(w^T \cdot \phi(x_i) - b) \geq 1 - \epsilon_i \text{ for } i = 1, \dots, N \quad \text{Eq.5-7}$$

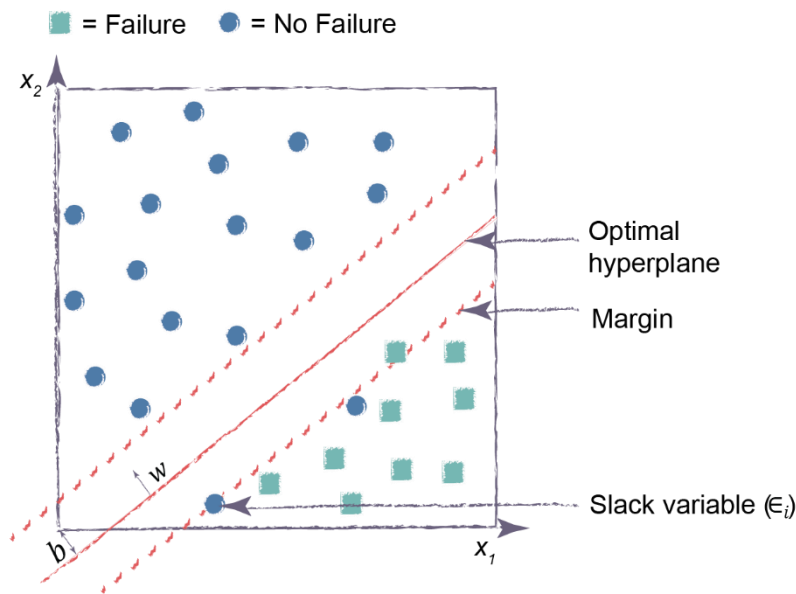


Figure 5-6 A simple one-dimensional illustration of an optimal linear hyperplane maximising the separation between two categories of data, with a single slack variable within the margin.

In one of the first pipe failure studies using SVM, Shirzad et al.(2014) used regression to predict the failure rate of pipes on two WDNs in Iran, Mashhad WDN with 580 km of pipe and Mahabad WDN with 140 km of pipe. The data set was limited to one year, with no further historical data, and only one pipe material was used for each WDN, asbestos cement and polyethylene, respectively. Comparing the results to an ANN, the authors found that ANN showed better statistical accuracy when predicting Polyethylene pipes (R^2 of 0.963 and 0.775 respectively), but similar results when comparing AC pipes (R^2 of 0.995 and 0.997 respectively). However, the authors suggested that the ANN did not generalise well since it was not consistent with the observed data, and the SVM was considered more suitable. Kutyłowska (2019) also used an SVM regression model on a WDN data set in Poland, including data collected between 2008-2014. The results showed a low relative error of between 4 and 14 % for the different pipe materials (cast iron, PVC, and polyethylene) and compared four kernel functions (linear, polynomial, sigmoidal and radial), finding Failure frequency is predicted well employing a linear kernel for all materials.

Robles-Velasco et al. (2020) compared a logistic regression and an SVM probabilistic classifier to predict the likelihood of failure in a medium-sized WDN of 3,800 km of pipes in Spain, recording failures between 2012-2018, using sampling techniques to balance the data, k -fold cross-validation for tuning, and pipe-intrinsic data and pipe pressure, but without the inclusion of any environmental variables. The study revealed that a global pipe model performed well (better than many materials specific models) and that the logistic regression showed similar results to the SVM, AUC 0.873 and 0.872, respectively. The logistic model prevented 34.09 % of all failures by replacing 3.16 % of the WDN, whilst the SVM prevented 29.52 % of failures by replacing 3.84 % of the WDN. The study managed data imbalance by under-sampling and removing data from the majority class, which can be a limitation since some information may have been lost. Giraldo-González and Rodríguez (2020) also used an SVM probabilistic classifier, comparing it against ANN, Bayes, and Gradient Boosting Trees (GBT) for asbestos cement and PVC pipes using age, diameter and length for a medium-sized WDN (1,819 km) on data collected between 2012-2018. The results found SVM to perform better on asbestos cement pipes over PVC pipes; AUC 0.991 and 0.795, respectively. Both Bayesian models and GBT outperformed the SVM. SVM performs poorly on datasets with noise and overlapping classes (common in pipe failure datasets) and is perhaps why they compare poorly to other machine learning models. The authors identified that SVM is useful for handling outliers by allowing misclassifications, letting some data fall within the margin to obtain better results, but limiting their use in the model (Giraldo-González and Rodríguez, 2020) via a slack variable ϵ_i applied to each of these data points (see Figure 5-6). Since SVM assumes the data is within a standard range, it is important to normalise the covariates.

5.4.3.5 Tree Models

Decision tree models are simple and can be used for regression and classification, and often as probabilistic classifiers (Robles-Velasco et al., 2021). Tree-based models have commonly been overlooked because they quickly

become complex and prone to overfitting and high variance (more so than other machine learning models). However, ensemble models significantly improve accuracy (Hastie et al., 2009) but are computationally expensive. With computational advances and quicker processing speeds, ensemble methods are more common and have shown better accuracy, even compared to several other data-driven pipe failure models (Chen et al., 2017; Giraldo-González and Rodríguez, 2020). Tree ensemble models are simple in their approach and do not require mathematical pre-processing steps, and for this reason, they are an attractive machine learning models that is easily accessible (Winkler et al., 2018).

Ensemble methods are best described in two parts, the decision tree, and the ensemble method. A decision tree T partitions data into disjoint regions R_j through recursive partitioning along the axis. The regions are split to minimise prediction errors until the size of the tree reaches the terminal nodes J and stops based on a stopping criterion (the Gini index is commonly used); the process is visualised in Figure 5-7. A decision tree is formally described by Hastie et al. (2009) as:

$$\hat{f}(x) = \sum_{j=1}^J c_R I(x \in R_j) \quad \text{Eq.5-8}$$

Where I is an indicator function, equal to 1 if the condition is true (failure) or 0 otherwise (non-failure). A simple model like a constant c_R is applied to each partitioned region that determines the probability in that region.

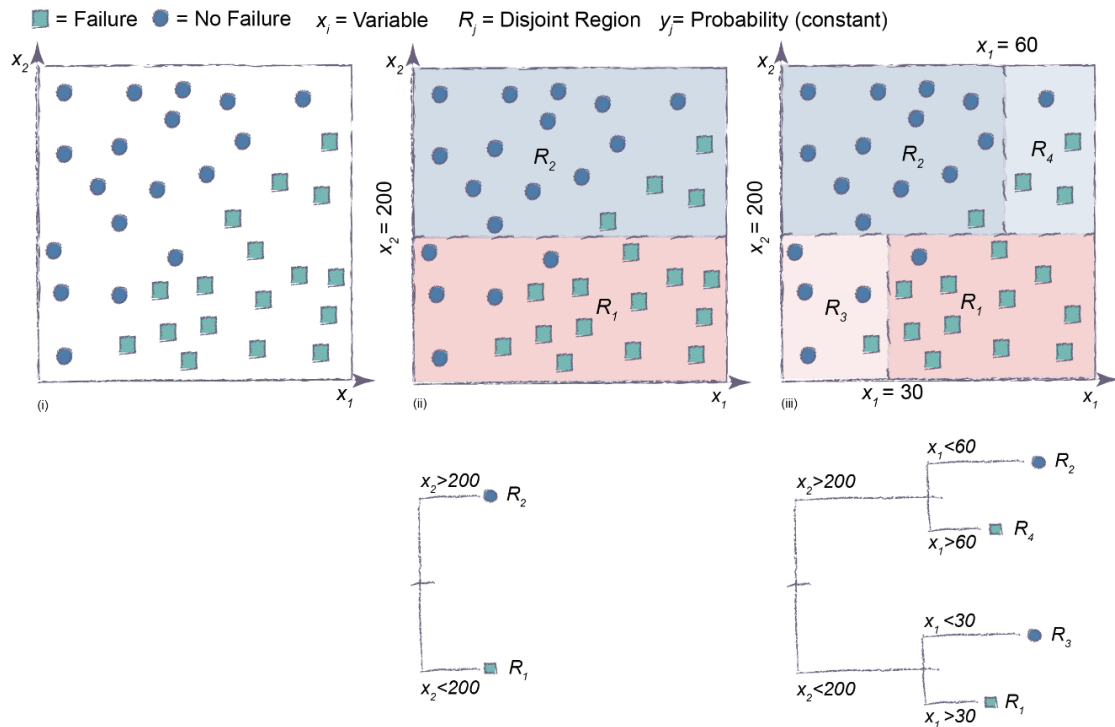


Figure 5-7 An example of data partitioning by a classification decision tree. (i) shows the two-dimensional data space. (ii) shows the first condition for splitting the data by variable x_2 at 200, where disjoint region R_1 is ≤ 200 and disjoint region R_2 > 200 . This creates two additional nodes for further conditions. (iii) shows the second condition for splitting data by variable x_1 at both 30 and 60 to create two more disjoint region R_3 and R_4 (adapted from Soguero-Ruiz et al. (2020)).

The ensemble method fits several trees, improving weak classifiers through bagging or boosting and averaging the results. Bagging decreases the variance by generating additional data from the training set, producing multiple original data sets. Bagging is illustrated in Figure 5-8, such that:

$$\hat{f}_{avg}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^b(x) \quad \text{Eq.5-9}$$

Where $\hat{f}^b(x)$ are the independent classifiers from B_i separate training sets are run in parallel, and the averaged of the trees is returned (Winkler et al., 2018).

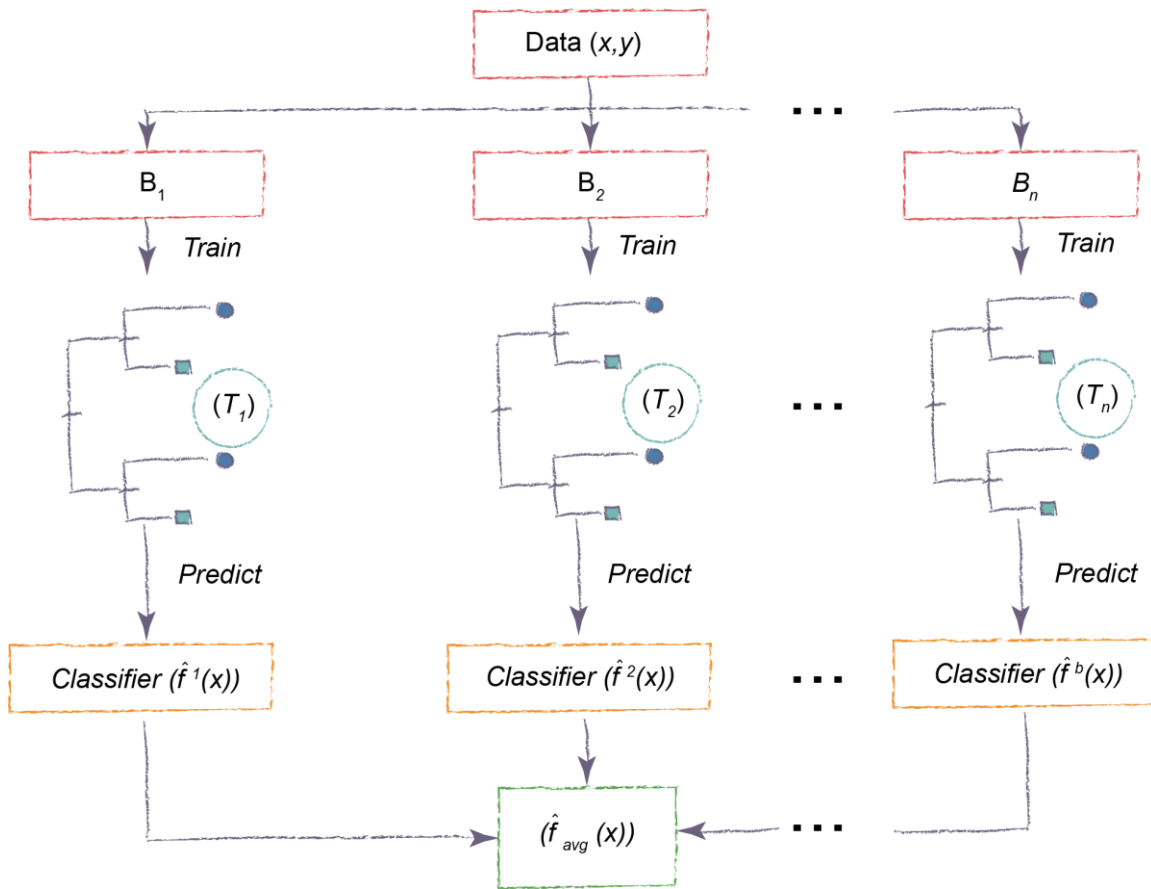


Figure 5-8 Depiction of the bagging process, showing the parallel trees returning classifiers averaged for the final model.

Boosting is conceptionally like bagging but uses the residuals of the previous tree to improve learning (Hastie et al., 2009). Taking a regression model as an example, gradient boosting aims to minimise the loss function in the existing collection of trees by adding, at each step, another tree that best reduces the loss function. The loss function is the residual of the response minus the fitted probability mean (Elith et al., 2008). The final regression gradient boosting model is depicted in Figure 5-9 and can be expressed as (Hastie et al., 2009):

$$\hat{f}_M(x) = \sum_{m=1}^M T(x; \gamma_m), \quad \text{Eq.5-10}$$

Where $\hat{f}_M(x)$ is the average of the tree values, M is the boosting iteration, T is a tree, and x is the multivariate argument characterised by a set of parameters γ .

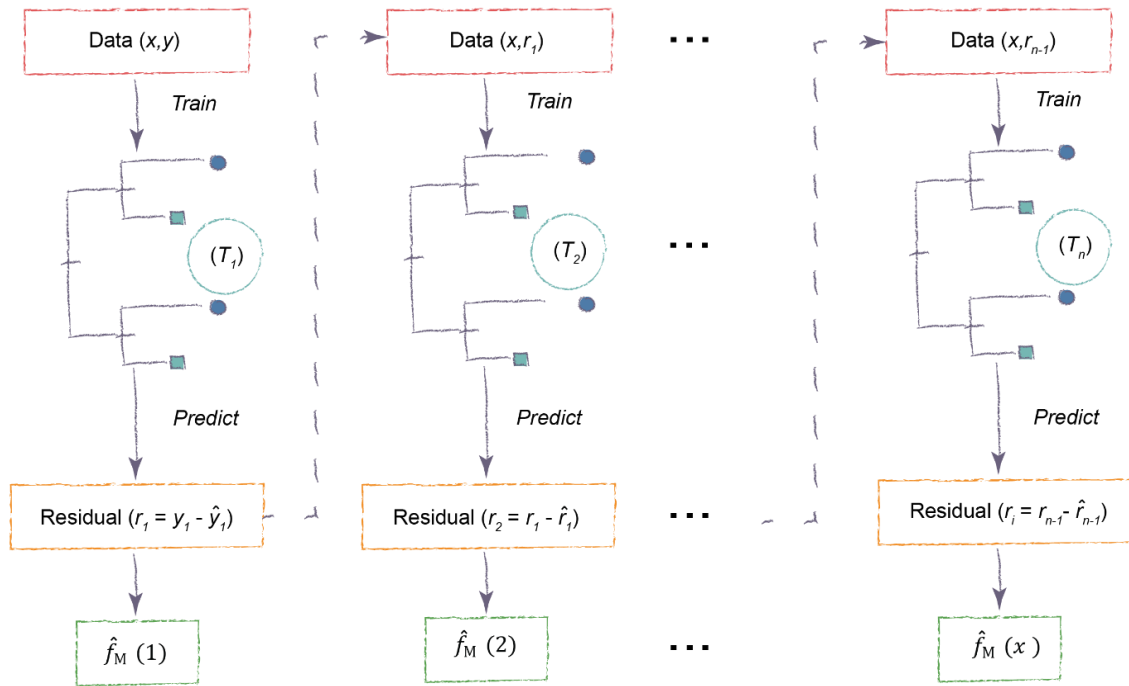


Figure 5-9 Depiction of Boosting, like bagging, but using the residuals of the previous tree to improve the learning. The final model is the average of all trees.

Winkler et al. (2018) created one of the first models using a variety of ensemble methods, applying the models to a medium-sized WDN in Austria of 851 km of individual pipes (39,637 pipes), with a failure record of > 30 years. The predictions were made over a long-time frame (five years and ten years) and compared RF to decision trees, namely Adaboost and RUSboost. The results found the RUSboost boosting method to be marginally better than the RF bagging method (AUC of 0.93 and 0.92, respectively). However, all models showed excellent results, partly due to dividing the pipes into street sections (having more equal lengths of pipe), dealing with the imbalanced data for all models through sampling methods, and using stratified sampling to represent the materials in both train and test datasets. Snider and McBean (2018) proposed using xgboost, an extreme gradient boosting model not previously used in literature, to predict time-to-failure for a 16,866 ductile iron water mains data set, recorded between 1960 to 2017. The xgboost model is faster than other gradient boosting models,

increasing its appeal in industry, yet the model requires extensive pre-processing of data, with categorical data requiring one-hot-encoding. The variables used in this study included the year of construction, district, pipe material, length, diameter, soil, lining, and cathodic protection variables. The xgboost was compared to a RF and ANN model using RMSE (test RMSE of 5.81, 5.90 and 7.32 respectively) and found the xgboost to have a 1.2 % improvement over the RF model and a 25.9 % improvement over the ANN model. In this study, the variable importance heavily leaned towards the previous failure date, whilst soils, district, and cathodic protection had limited importance. Overall the authors found xgboost to be a reliable option, following with further research (Snider and McBean, 2019). Chen et al. (2019) compared many models, including GLM, GAM, RF, GBT and Generalised Linear Mixed Models (GLMM), on pipes divided by road network. The authors concluded that the GBT model performs best, returning the lowest Brier score (a more accurate measure for probabilistic predictions). The main limitations here are the monthly prediction interval and limited access to variables correlated to pipe failure. GBT models are useful, yet they can be computationally expensive and prone to overfitting if the hyperparameters are not tuned correctly. No calibration curve is provided. Giraldo-González and Rodríguez (2020) compared SVM, ANN, Bayes and GBT for a medium-sized WDN over a long prediction interval (five years). Predicting the probability of individual pipe failures, the authors found the GBT to have the best performance (AUC 0.998 for AC pipes and 0.990 for PVC pipes), although again, all models showed excellent results. The limitation here is that the authors present the probability of failure and yet do not provide a calibration curve to show how well the probabilities are observed to the outcomes in the sample test data.

The main limitation in most GBT studies is a lack of information on the model's performance, particularly the absence of a calibration plot or the brier score to show the accuracy of the probabilities, which is helpful for models that are probabilistic classifiers. Further issues include the difficulty tuning

hyperparameters and the risk of overfitting due to the additive process and the regularization criteria used to control complexity (Wei et al., 2018).

5.4.3.6 Comments

Advancing into machine learning for pipe failure modelling is an obvious choice since these approaches are capable of modelling complex relationships between variables and compliment the increased data collection over the past decade. Machine learning has been shown to improve predictive accuracy compared to deterministic and probabilistic approaches (Konstantinou and Stoianov, 2020), although some studies show similar accuracy (Chen et al., 2019; Robles-Velasco et al., 2020), and when methodologies include large groups of pipes, deterministic models show comparable accuracy, as shown by Giraldo-González and Rodríguez (2020). Machine learning models can be used for regression, classification, or as probabilistic classifiers.

Supervised machine learning models are preferred since they perform well on structured data and are interpretable through variable importance and partial plots, amongst other widely used accuracy metrics. The predictive capability of machine learning models works best with larger datasets. However, there is a greater need for advanced pre-processing methods, such as data transformation, encoding categorical data through one-hot-encoding or dummy coding and standardisation (e.g., maximum, and minimum values), which are essential since the models are often sensitive to outliers. Pre-processing methods can be extensive and can consume some 60 to 80 % of the effort (Kahn, 2021). Cross-validation and hyperparameter tuning are essential for reducing overfitting and bias. Machine learning is a complex mathematical framework that could deter its use within industry.

Clustering is a straightforward process and is used to identify spatial relationships between failures. These relationships were successfully used by De Oliveira et al. (2011a) to target areas for investigation. However, this approach is backwards-looking and cannot predict future failures. Nevertheless, spatial

relationships are essential, and Chen and Guikema (2020) used the information as a variable in the dataset, suggesting beneficial results.

ANN is widely used for its flexibility, firstly, the algorithm automatically adjusts the connection weights to reflect non-linearities in the data, and secondly, the ANN can be developed using multiple training algorithms. This flexibility continues through extensions of alternative algorithms to help improve results, such as fuzzy logic used by Christodoulou and Deligianni (2010) and the neuro-fuzzy inference system used by Tabesh et al. (2009). The ANN framework can generalise well but is also prone to overfitting if not tuned correctly. Some studies have reported poor generalisation (Shirzad et al., 2014), but this may be the result of limited data typically seen in pipe failure data sets, especially considering ANN ordinarily performs better on high dimensional unstructured data. The complexity of the model structure is problematic for two reasons, firstly the optimum structure of an ANN is identified a priori (i.e., inputs, hidden layers and data transformation within the activation units), such that a time-consuming trial and error approach is typical, and knowledge of the weight terms or bias is generally unknown, which limits interpretability (Xu et al., 2011).

EPR models provide transparent and well-structured relationships from a simple, easy-to-use formula that water companies can apply. Unlike other machine learning models, EPR has few parameters to tune, can include expert knowledge and is reasonably fast. However, EPR, like other machine learning models, is computationally expensive and prone to overfitting, resulting in poor accuracy and overestimating failures in pipes with no recorded failures (Berardi et al., 2008). Another major disadvantage is that only a single output can be achieved since the model is regression, and therefore works best on groups of pipes to return the failure rates (Giraldo-González and Rodríguez, 2020).

SVM is a flexible model that can be a regression or a probabilistic classifier and improves linear models by modelling linear relationships with a single hyperplane. Non-linear relationships can be implicitly modelled using a kernel function, which can be difficult. This is advantageous since the approach is efficient with high

dimensional data. The algorithm fails to cope with noisy data where the clear separation between classifications is not easily defined; poor quality has been observed in pipe failure data (Debón et al., 2010; Kabir et al., 2016; Tang et al., 2019). Considering this, SVM is often outperformed by different machine learning models (Giraldo-González and Rodríguez, 2020) and also simplistic mathematical frameworks like the logistic regression model (Robles-Velasco et al., 2020), which is perhaps why the model is not often used.

Ensemble tree methods consistently outperform other machine learning models, as found in several comparison studies (Chen et al., 2019; Giraldo-González and Rodríguez, 2020). Bagging and boosting are two approaches commonly used in literature, and a comparison of different methods by Winkler et al. (2018) has shown that gradient boosting is statistically more accurate. All the methods in this study showed good results predicting the probability of failure, although accuracy measures, such as calibration plots or brier scores, were not presented. The methodology for many ensemble studies has included stratification when dividing test and train data sets and predicting over long periods, typically five years, adding to the improved accuracy (Giraldo-González and Rodríguez, 2020; Winkler et al., 2018). Another advantage of ensemble methods is that data standardisation is not required, making the pre-processing slightly easier. Table 5-3 shows a summary of the main machine learning models discussed.

Table 5-3 Summary of the main machine learning pipe failure models.

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Christodoulou et al. (2003)	ANN Fuzzy logic	365 km	1982 to 2002	AC, CI, DI	Pipe-intrinsic Environmental	Pipe groups	Number of failures (interarrival time)
Berardi et al. (2008)	EPR	172 km (3,669 pipes)	1986 to 1999	-	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Tabesh et al. (2009)	ANN ANFIS	580 km		AC, CI	Pipe-intrinsic Operational	-	Failure rate
Christodoulou and Deligianni (2010)	ANN Fuzzy logic	365km 795 km	10 years	CI, Plastic, ST	Pipe-intrinsic	individual pipes	Risk of failure
Jafar et al. (2010)	ANN	162 km (4,862 pipes)	1991 to 2004	Cement, CI, DI, Plastic	Pipe-intrinsic Environmental Operational	Individual pipes	Number of failures
Nishiyama and Fillion (2013)	ANN	670 km	1998 to 2011	CI	Pipe-intrinsic Environmental	Pipe groups	Number of Failures
Asnaashari et al. (2013)	ANN	784 km	1959 to 2004	CI, Concrete, DI, AC, PVC	Pipe-intrinsic Environmental	Pipe groups	Failure rate

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Harvey et al. (2014)	ANN	1,021 km (6,346 pipes)	1962 to 2005	AC, CI, DI, PVC	Pipe-intrinsic Environmental	Individual pipes	Time-to-failure (interarrival time)
Laucelli et al. (2014)	EPR	679 km (6,879 pipes)	1962 to 2003	CI	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Shirzad et al. (2014)	ANN SVR	580 km 140 km	One year	AC, PE	Pipe-intrinsic Operational	Pipe groups	Failure rate
Farmani et al. (2017)	EPR	300.63 km (7987 pipes)	-	CI	Pipe-intrinsic Environmental	Pipe groups	Number of failures
Kakoudakis et al. (2017)	EPR	-	-	AC, CI, DI, PE, PVC	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Kakoudakis et al. (2018)	ANN EPR	647 km (18,872 pipes)	2003 to 2013	CI	Pipe-intrinsic Environmental	Pipe groups	Failure rate
Winkler et al. (2018)	Adaboost RUSboost RF GBT	851 km (39,637)	> 30 years	CI, DI, PE, ST	Pipe-intrinsic Operational	Individual pipes	Probability of failure

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
Snider and McBean (2018)	GBT (xgboost) RF ANN	3,042 km	1960 to 2017	Cement, CI, DI, PVC	Pipe-intrinsic Environmental Operational	Individual pipes	Time-to-failure
Sattar et al. (2019)	Extreme Learning Machine (ELM) ANN SVM Non-linear regression	> 1000 km (6,342 pipes)	from 1962	AC, CI, DI	Pipe-intrinsic Environmental	Individual pipes	Time-to-failure
Kutyłowska (2019)	SVM	17 km 14 km	2008 to 2014	CI, PE, PVC	Pipe-intrinsic Operational	Individual pipes	Failure rate
Chen et al. (2019)	GAM RF DT GBT GLMM	-	2010 - 2014	-	Environmental Census	Individual pipes (Road segment) Pipe group (census track)	Probability of failure
Giraldo-González and Rodríguez (2020)	Linear, Poisson, and EPR GBT, Bayes, SVM,	652 km (20,793 pipes)	2015 to 2017	AC, CI, DI, PE, PVC	Pipe-intrinsic Environmental Operational	Pipe groups Individual pipes	Failure rates Probability of failure

Author	Model ¹	Network Size	Failure History	Pipe Materials ²	Variables	Spatial Level	Response
	ANN						
Robles-Velasco et al. (2020)	SVM Linear	3,800 km of	2012 to 2018	Cements, metal, plastics	Pipe-intrinsic Operational	Individual pipes	Probability of failure
Chen and Guikema (2020)	GLM GAM RF GBT	680 km (12,092 pipes)	2008 to 2017	CI, other	Census Pipe-intrinsic Environmental Operational	Individual pipes	Number of failures
Aslani et al. (2021)	RF GBT Multivariate adaptive regression splines ANN	76,000 pipes	2015 to 2020	CI, DI, PE, PVC	Pipe-intrinsic Environmental Operational	Individual pipes	Failure rate

- = Unavailable information

1 ANN = Artificial Neural Network, DT = Decision Tree, EPR = Evolutionary Polynomial Regression, GAM = Generalized Additive Model, GBT = Gradient Boosting Tree, GLM = Generalised Linear Model, GLMM = Generalised Linear Mixed Effect Model, RF = Random Forest, SVM = Support Vector Machine

2 AC = Asbestos Cement, CI = Cast Iron, DI = Ductile Iron, HY = Hyprescon, PE = Polyethylene, PVC = Polyvinyl Chloride, ST = Steel

5.5 Summary and Statistical Model Decision Aid

This review has focussed on providing an overview of the standard methods and models used to predict pipe failures. The evolution of pipe failure models is clear, from simplistic single variate regression models at a network scale to the use of more complex machine learning, able to accommodate large multivariate datasets and finer individual pipe scale methods. Each category of statistical model has advantages and disadvantages, detailed in Table 5-4.

Table 5-4 Advantages and disadvantages of deterministic, probabilistic and machine learning methods.

Category	Advantages	Disadvantages
Deterministic	<ul style="list-style-type: none">▪ Simplistic mathematical frameworks.▪ Easy to implement within industry.▪ Easy to develop and quick to run.▪ Interpretable results.▪ Easy to develop.▪ Computationally efficient.	<ul style="list-style-type: none">▪ It cannot account for randomness.▪ Does not easily handle large complex data sets.▪ Cannot be used to return the probability of failure.▪ Not accurate for individual pipe models.▪ Not flexible (cannot be tuned)
Probabilistic	<ul style="list-style-type: none">▪ Survival analysis can predict different outcomes and cope with left-truncated data.▪ Can account for randomness.▪ The probability of failure is more accurate for individual pipe models.▪ Bayes models can incorporate expert knowledge.	<ul style="list-style-type: none">▪ Some models have complex mathematical frameworks.▪ Not easy for water companies to implement as it would require expert, specialised knowledge.▪ Survival analysis requires extensive failures records.▪ The probability of failure is not always required.
Machine Learning	<ul style="list-style-type: none">▪ Can describe complex relationships for large datasets.	<ul style="list-style-type: none">▪ Complex mathematical frameworks.▪ Not easy for water companies to implement.

Category	Advantages	Disadvantages
	<ul style="list-style-type: none">▪ Generally, it improves accuracy for individual pipe models.▪ Can be used to return failure rate, time-to-failure, and as probabilistic classifiers.▪ Are flexible and can be tuned using hyperparameters.	<ul style="list-style-type: none">▪ It would require expert knowledge.▪ Computationally expensive to run and during pre-processing.▪ Prone to overfitting.▪ Results are harder to interpret.

Static variables are fundamental to pipe failure models, and throughout the literature pipe intrinsic factors such as pipe material, pipe age, pipe length, and the number of failures are among the most significant. The temporal scale of predicting pipe failures has moved towards predicting shorter time scales, motivated by the need to assist operational management. Shifting towards predicting failures over shorter time intervals can use time-dynamic variables that capture the effects of seasonal or annual variations (and potentially future climatic variations). However, modelling time-dependent dynamic variables is more complex, especially for models that attempt to predict response for individual pipes, because pipe failures are rare and result in imbalanced data. In this scenario, the assumptions of the Poisson distribution are not met; therefore, predicting the number of failures or failure rate has resulted in poor accuracy. Predicting the probability of failure or the time-to-failure is reported as more useful for individual pipes and is often enough information for decision-makers. Zero-inflated models may be used, but these have not necessarily proved to be more accurate. Alternatively, models should group pipes for greater accuracy if short prediction intervals are necessary. The Poisson distribution can return the number of failures or failure rate if the data is such that enough failures are captured by pipe group for statistical significance. However, in such instances, pipes that regularly fail may also be included in a group with pipes that never fail, distorting the true failure rate for many pipes. Therefore, grouping must be

considered carefully, with some authors choosing census level groups to provide more accurate predictions. Concisely, time-dependent dynamic variables can be included in pipe failure models, though the statistical analysis may prove to be more challenging.

The most statistically accurate temporal scale for model accuracy on individual pipes remains long term prediction intervals, annual or greater, and in many recent machine learning efforts, individual pipes are predicted over five or more years (Harvey et al., 2014; Nishiyama and Fillion, 2013; Robles-Velasco et al., 2020; Snider and McBean, 2019). Long-term prediction intervals show favourable accuracy where enough failures accumulate for statistical significance but at a loss of inter-annual information. Survival analysis lends itself well to long term planning because there are options to predict interarrival times of failures. Advances in survival analysis have shown valuable results based on the Weibull distribution used for time-to-first failure and the exponential distribution for subsequent failures. Typically, a long pipe failure data set is required. Alternatively, Bayesian models return the probability of failure and accommodate explicit expert domain knowledge using prior distributions. The graphical nature of Bayesian models provides a useful means of viewing the interactions between failures and the variables. If the objective is to analyse the results, then Bayesian networks are appropriate due to their interpretability. Yet, a lack of failure data means the influence of prior information is greater, resulting in bias introduced from opinions.

Machine learning models are not strictly probabilistic but can be termed probabilistic classifiers and have also proved to be useful when predicting the probability of failure for individual pipes over long periods (Winkler et al., 2018). Supervised models are the obvious choice for the structured nature of pipe failure datasets, and comparing machine learning models, the gradient boosting model has shown great promise in terms of accuracy (Giraldo-González and Rodríguez, 2020), yet SVM and ANN have also shown useful results (Robles-Velasco et al., 2020). There is a level of complexity in machine learning that can deter its use in

industry, and therefore very few models have transferred into industry working models. The complexity is associated with pre-processing, often, extensive datasets since the models are more sensitive to outliers; one-hot-encoding or dummy coding and standardisation (e.g., maximum, and minimum values) are two examples. Tuning hyperparameters using cross-validation can be difficult but is essential to avoid overfitting. However, machine learning models are still subject to the same difficulties when considering the spatiotemporal scale and the limited number of failures observed in pipe failure datasets. Classification models predicting probability can use various sampling techniques (e.g., under or oversampling) to balance the data, yet there are limitations to these techniques, and the results are not always favourable, nor do they necessarily improve the model accuracy (Fan et al., 2022).

Choosing the correct model is a complex process and can directly impact the usefulness of the response for decision-makers. No single model is superior, and so the choice of model should be one that carefully considers four key factors, 1) the type of variables available, 2) the time interval for management interventions, 3) the spatial level of the model (e.g., pipes or grouped pipes), and 4) the response type and level of inference required. Practitioners may find it useful to understand the most appropriate model; thus, Figure 5-10 shows a flow diagram devised to map these general assumptions, leading to a model selection that is likely to be the most useful in different decision-making contexts. The model shows five steps, with the first four steps working through the key factors and the final step suggesting some suitable statistical models.

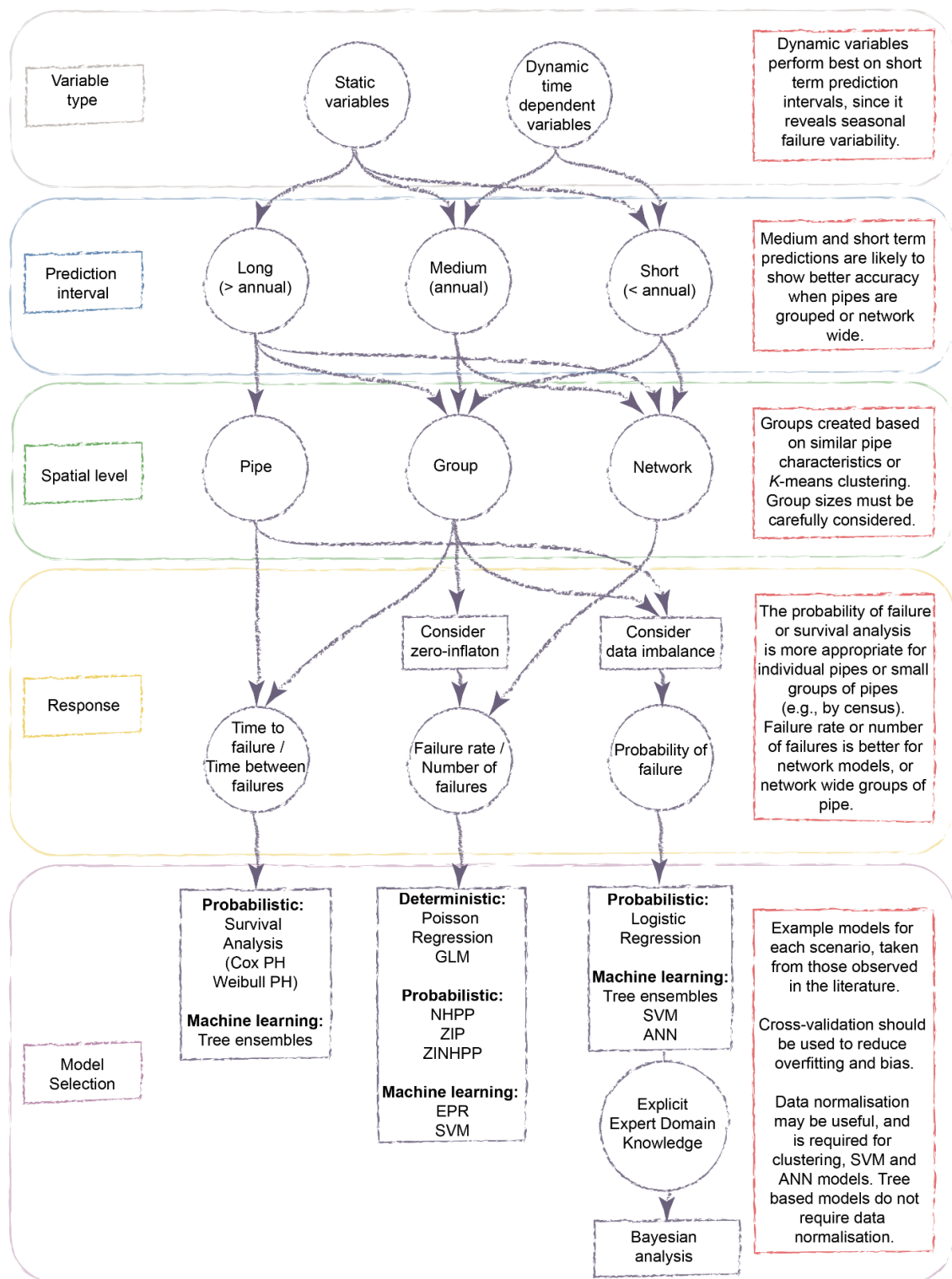


Figure 5-10 Prediction model decision aid to determine the most useful approach in different decision-making contexts, considering four important key factors, 1) the type of variables available, 2) the time interval for management interventions,

3) the spatial level of the model, and 4) the response type and level of inference required.

5.6 Conclusions

From the literature, several general conclusions are deduced:

- There is much discussion on the benefits of more complex models, such as Bayesian and machine learning; however, these models are still subject to poor data and limited data. Considering this, simple models are often useful and more intuitive.
- Early failure models separated pipe materials for more accurate results due to their unique behaviour within environments. Yet, recent literature has shown improved results for global models, where failure can be accumulated to avoid non-convergence due to too few failures. Using stratified sampling, the global model ensures representative materials are found in both the training and test data sets. This benefits from developing a single model instead of multiple models for various materials, saving time and providing a denser dataset for modelling.
- Data quality is generally poor, yet only a handful of studies explicitly identify data quality as a limiting issue. Furthermore, many studies do not include clarified assumptions or study limitations; for example, the approach to pre-processing the bursts data is often unknown, and yet there can be several issues in locating pipe failures accurately along with a network. These issues should be detailed to provide a better understanding of the limitations.
- It has been acknowledged that failure records collected by water companies are typically short (Yamijala et al., 2009; Chen et al., 2019). From the studies used here (noting this is not an exhaustive list and that not all literature had pipe failure record length), some records are as short as three years; however, approximately 60 % had failure records greater than ten years, and 35 % greater than 30 years. The studies with longer failure records report more useful results; however, it is unclear if the

length of failure records or the length of the prediction interval improves the results, perhaps a combination of both.

- Water utilities should endeavour to collect regular data in such a manner to facilitate long-term studies and modelling efforts.
- Some studies do not demonstrate how the reported model can be transferred into a practical tool and utilised by water companies. Furthermore, most of the models were not implemented in ways whereby the outcomes could be easily integrated into a GIS framework for visualisation and further spatial analysis.
- Other than clustering algorithms that are not strictly failure models, there is a limited scope to address the spatiotemporal relationship between pipe failures. More research could be usefully conducted in this area, especially incorporating the use of cluster analysis.
- Limited studies have been undertaken on large networks, and many studies focus efforts in urban areas, typically cities. More studies could be completed using rural areas or a combination of both.
- Few studies use a comprehensive list of pipe-intrinsic, environmental, and operational covariates, which shows a lack of available data rather than shortfalls in the studies themselves.
- Few studies include data from decommissioned pipes, which would increase the dataset and number of failures to train the models.

The following is a list of recommendations related to improving the practice of pipe failure modelling for future methodologies:

- Data is essential to modelling failures correctly. At a minimum, pipe-intrinsic data should be collected. Further research on census data and socio-economic factors could provide interesting insight.
- Data collection and management protocols are needed to improve data quality and quantity.
- Additional research is required on effective data collection and information extraction from existing data.

- Data collection and temporal and spatial scales should be tied to the model that is used. Further research is needed to establish appropriate pipe groupings that improves accuracy whilst describing the influences of more localised conditions.

5.7 Author Contributions

Neal Barton: Conceptualisation, Methodology, Investigation, Writing – Original Draft, Writing – Review and Editing, Visualisation and Project administration. Stephen Hallett: Writing – Review and Editing, Supervision, Funding Acquisition. Simon Jude: Review, Supervision, Funding Acquisition. Trung Hieu Tran: Review and Editing.

5.8 References

- Alvisi, S., Franchini, M., 2010. Comparative analysis of two probabilistic pipe breakage models applied to a real water distribution system. *Civ. Eng. Environ. Syst.* 27, 1–22. <https://doi.org/10.1080/10286600802224064>
- Andreou, S.A., Marks, D.H., Clark, R.M., 1987a. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. *Adv. Water Resour.* 10, 2–10. [https://doi.org/10.1016/0309-1708\(87\)90002-9](https://doi.org/10.1016/0309-1708(87)90002-9)
- Andreou, S.A., Marks, D.H., Clark, R.M., 1987b. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. *Adv. Water Resour.* 10, 2–10. [https://doi.org/10.1016/0309-1708\(87\)90002-9](https://doi.org/10.1016/0309-1708(87)90002-9)
- Aslani, B., Mohebbi, S., Axthelm, H., 2021. Predictive analytics for water main breaks using spatiotemporal data. *Urban Water J.* 00, 1–16. <https://doi.org/10.1080/1573062X.2021.1893363>

- Asnaashari, A., McBean, E.A., Gharabaghi, B., Tutt, D., 2013. Forecasting watermain failure using artificial neural network modelling. *Can. Water Resour. J.* 38, 24–33. <https://doi.org/10.1080/07011784.2013.774153>
- Asnaashari, A., McBean, E.A., Shahrour, I., Gharabaghi, B., 2009. Prediction of watermain failure frequencies using multiple and Poisson regression. *Water Supply* 9, 9–19. <https://doi.org/10.2166/ws.2009.020>
- Barton, N.A., Farewell, T.S., Hallett, S.H., 2020. Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *npj Clean Water* 3, 31. <https://doi.org/10.1038/s41545-020-0077-3>
- Barton, N.A., Farewell, T.S., Hallett, S.H., Acland, T.F., 2019. Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Res.* 164, 114926. <https://doi.org/10.1016/j.watres.2019.114926>
- Barton, N.A., Hallett, S.H., Jude, S.R., 2021. The challenges of predicting pipe failures in clean water networks: a view from current practice. *Water Supply*. <https://doi.org/10.2166/ws.2021.255>
- Berardi, L., Giustolisi, O., Kapelan, Z., Savic, D.A., 2008. Development of pipe deterioration models for water distribution systems using EPR. *J. Hydroinformatics* 10, 113–126. <https://doi.org/10.2166/hydro.2008.012>
- Chen, T.Y.-J., Beekman, J.A., David Guikema, S., Shashaani, S., 2019. Statistical Modeling in Absence of System Specific Data: Exploratory Empirical Analysis for Prediction of Water Main Breaks. *J. Infrastruct. Syst.* 25, 04019009. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000482](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000482)
- Chen, T.Y., Beekman, J.A., Guikema, S.D., 2017. Drinking Water Distribution Systems Asset Management: Statistical Modelling of Pipe Breaks, in: *Pipelines 2017*. ASCE, pp. 173–186.

- Chen, T.Y.J., Guikema, S.D., 2020. Prediction of water main failures with the spatial clustering of breaks. *Reliab. Eng. Syst. Saf.* 203, 107108. <https://doi.org/10.1016/j.ress.2020.107108>
- Christodoulou, S., Aslani, P., Vanrenterghem, A., 2003. A Risk Analysis Framework for Evaluating Structural Degradation of Water Mains in Urban Settings, Using Neurofuzzy Systems and Statistical Modeling Techniques, in: *World Water & Environmental Resources Congress 2003*. American Society of Civil Engineers, Reston, VA, pp. 1–9. [https://doi.org/10.1061/40685\(2003\)134](https://doi.org/10.1061/40685(2003)134)
- Christodoulou, S., Charalambous, B., Adamou, A., 2007. Managing the “repair or replace” dilemma on water leakages, in: *International Water Association’s (IWA) Third Specialty Conference on Water Loss Reduction (Water Loss 2007)*. Bucharest, Romania.
- Christodoulou, S., Deligianni, A., 2010. A Neurofuzzy Decision Framework for the Management of Water Distribution Networks. *Water Resour. Manag.* 24, 139–156. <https://doi.org/10.1007/s11269-009-9441-2>
- Clark, R.M., Stafford, C.L., Goodrich, J.A., 1982. Water distribution systems: a spatial and cost evaluation. *J. Water Resour. Plan. Manag.* 108, 243–256.
- Cox, D.R., 1972. Regression Models and Life-Tables. *J. R. Stat. Soc. Ser. B* 34, 187–220.
- De Oliveira, D.P., Garrett, J.H., Soibelman, L., 2011. A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage. *Adv. Eng. Informatics* 25, 380–389. <https://doi.org/10.1016/j.aei.2010.09.001>
- Deadman, C., 2010. *Strategic Asset Management: The quest for utility excellence*. Matador, Leicester, UK.

- Debón, A., Carrión, A., Cabrera, E., Solano, H., 2010. Comparing risk of failure models in water supply networks using ROC curves. *Reliab. Eng. Syst. Saf.* 95, 43–48. <https://doi.org/10.1016/j.ress.2009.07.004>
- Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numer. Math.* 1, 269–271. <https://doi.org/10.1007/BF01386390>
- Dridi, L., Mailhot, A., Parizeau, M., Villeneuve, J.P., 2005. A strategy for optimal replacement of water pipes integrating structural and hydraulic indicators based on a statistical water pipe break model, in: *Proceedings of the 8th International Conference on Computing and Control for the Water Industry*. Exeter, UK, pp. 65–70.
- Economou, T., Kapelan, Z., Bailey, T., 2008. A Zero-Inflated Bayesian Model for the Prediction of Water Pipe Bursts, in: *Water Distribution Systems Analysis 2008*. American Society of Civil Engineers, Reston, VA, pp. 1–11. [https://doi.org/10.1061/41024\(340\)61](https://doi.org/10.1061/41024(340)61)
- Economou, T., Kapelan, Z., Bailey, T.C., 2012. On the prediction of underground water pipe failures: zero inflation and pipe-specific effects. *J. Hydroinformatics* 14, 872–883. <https://doi.org/10.2166/hydro.2012.144>
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Fan, X., Wang, X., Zhang, X., ASCE Xiong (Bill) Yu, P.E.F., 2022. Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors. *Reliab. Eng. Syst. Saf.* 219, 108185. <https://doi.org/10.1016/j.ress.2021.108185>
- Farmani, R., Kakoudakis, K., Behzadian, K., Butler, D., 2017. Pipe Failure Prediction in Water Distribution Systems Considering Static and Dynamic Factors. *Procedia Eng.* 186, 117–126. <https://doi.org/10.1016/j.proeng.2017.03.217>

- Fitchett, J.C., Karadimitriou, K., West, Z., Hughes, D.M., 2020. Machine Learning for Pipe Condition Assessments. *J. Am. Water Works Assoc.* 112, 50–55. <https://doi.org/10.1002/awwa.1501>
- Francis, R.A., Guikema, S.D., Henneman, L., 2014. Bayesian Belief Networks for predicting drinking water distribution system pipe breaks. *Reliab. Eng. Syst. Saf.* 130, 1–11. <https://doi.org/10.1016/j.ress.2014.04.024>
- García, A.C., Solano, H., Gamiz, M.L., Debón, A., 2010. Evaluation of the Reliability of a Water Supply Network from Right-Censored and Left-Truncated Break Data. *Water Resour. Manag.* 24, 2917–2935. <https://doi.org/10.1007/s11269-010-9587-y>
- Getis, A., Ord, J.K., 1992. The Analysis of Spatial Association by Use of Distance Statistics. *Geogr. Anal.* 24, 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Giraldo-González, M.M., Rodríguez, J.P., 2020. Comparison of Statistical and Machine Learning Models for Pipe Failure Modeling in Water Distribution Networks. *Water* 12, 1153. <https://doi.org/10.3390/w12041153>
- Giustolisi, O., Savic, D.A., 2006. A symbolic data-driven technique based on evolutionary polynomial regression. *J. Hydroinformatics* 8, 207–222. <https://doi.org/10.2166/hydro.2006.020b>
- Goulter, I.C., Kazemi, A., 1988. Spatial and temporal groupings of water main pipe breakage in Winnipeg. *Can. J. Civ. Eng.* 15, 91–97. <https://doi.org/10.1139/l88-010>
- Harvey, R., McBean, E.A., Gharabaghi, B., 2014. Predicting the Timing of Water Main Failure Using Artificial Neural Networks. *J. Water Resour. Plan. Manag.* 140, 425–434. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000354](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000354)

- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics, Springer Series in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-84858-7>
- Jafar, R., Shahrour, I., Juran, I., 2010. Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. Math. Comput. Model. 51, 1170–1180. <https://doi.org/10.1016/j.mcm.2009.12.033>
- Kabir, G., Balek, N.B.C., Tesfamariam, S., 2018. Consequence-based framework for buried infrastructure systems: A Bayesian belief network model. Reliab. Eng. Syst. Saf. 180, 290–301. <https://doi.org/10.1016/j.ress.2018.07.037>
- Kabir, G., Tesfamariam, S., Loepky, J., Sadiq, R., 2016. Predicting water main failures: A Bayesian model updating approach. Knowledge-Based Syst. 110, 144–156. <https://doi.org/10.1016/j.knosys.2016.07.024>
- Kahn, C., 2021. The Value Proposition for Likelihood-of-Failure Modeling. J. AWWA 113, 30–38. <https://doi.org/10.1002/awwa.1649>
- Kakoudakis, K., Behzadian, K., Farmani, R., Butler, D., 2017. Pipeline failure prediction in water distribution networks using evolutionary polynomial regression combined with K -means clustering. Urban Water J. 14, 737–742. <https://doi.org/10.1080/1573062X.2016.1253755>
- Kakoudakis, K., Farmani, R., Butler, D., 2018. Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. J. Hydroinformatics 20, 1191–1200. <https://doi.org/10.2166/hydro.2018.152>
- Kalogirou, S.A., 2004. Neural Network Modeling of Energy Systems, in: Encyclopedia of Energy. Elsevier, pp. 291–299. <https://doi.org/10.1016/B0-12-176480-X/00536-2>

- Kettler, A.J., Goulter, I.C., 1985. An analysis of pipe breakage in urban water distribution networks. *Can. J. Civ. Eng.* 12, 286–293. <https://doi.org/10.1139/l85-030>
- Kimutai, E., Betrie, G., Brander, R., Sadiq, R., Tesfamariam, S., 2015. Comparison of Statistical Models for Predicting Pipe Failures: Illustrative Example with the City of Calgary Water Main Failure. *J. Pipeline Syst. Eng. Pract.* 6, 04015005. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000196](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000196)
- Kleiner, Y., Nafi, A., Rajani, B., 2010. Planning renewal of water mains while considering deterioration, economies of scale and adjacent infrastructure. *Water Supply* 10, 897–906. <https://doi.org/10.2166/ws.2010.571>
- Kleiner, Y., Rajani, B., 2012. Comparison of four models to rank failure likelihood of individual pipes. *J. Hydroinformatics* 14, 659–681. <https://doi.org/10.2166/hydro.2011.029>
- Kleiner, Y., Rajani, B., 2010. I-WARP: Individual Water Main Renewal Planner. *Drink. Water Eng. Sci.* 3, 71–77. <https://doi.org/10.5194/dwes-3-71-2010>
- Kleiner, Y., Rajani, B., 2002. Forecasting Variations and Trends in Water-Main Breaks. *J. Infrastruct. Syst.* 8, 122–131. [https://doi.org/10.1061/\(ASCE\)1076-0342\(2002\)8:4\(122\)](https://doi.org/10.1061/(ASCE)1076-0342(2002)8:4(122))
- Kleiner, Y., Rajani, B., 2001. Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water* 3, 131–150. [https://doi.org/10.1016/S1462-0758\(01\)00033-4](https://doi.org/10.1016/S1462-0758(01)00033-4)
- Kleiner, Y., Rajani, B., 2000. Considering time-dependent factors in the statistical prediction of water main breaks, in: *American Water Works Association Infrastructure Conference (AWWA 2000)*. Seattle, USA, pp. 1–12.
- Kleiner, Y., Sadiq, R., Rajani, B., 2004. Modeling failure risk in buried pipes using fuzzy Markov deterioration process. *Proc. ASCE Pipeline Div. Spec.*

- Congr. - Pipeline Eng. Constr. 40745, 67–78.
[https://doi.org/10.1061/40745\(146\)7](https://doi.org/10.1061/40745(146)7)
- Konstantinou, C., Stoianov, I., 2020. A comparative study of statistical and machine learning methods to infer causes of pipe breaks in water supply networks. Urban Water J. 17, 534–548.
<https://doi.org/10.1080/1573062X.2020.1800758>
- Kutyłowska, M., 2019. Forecasting failure rate of water pipes. Water Supply 19, 264–273. <https://doi.org/10.2166/ws.2018.078>
- Kutyłowska, M., 2015. Modelling of Failure Rate of Water-pipe Networks. Period. Polytech. Civ. Eng. 59, 37–43. <https://doi.org/10.3311/PPci.7541>
- Laucelli, D., Rajani, B., Kleiner, Y., Giustolisi, O., 2014. Study on relationships between climate-related covariates and pipe bursts using evolutionary-based modelling. J. Hydroinformatics 16, 743–757.
<https://doi.org/10.2166/hydro.2013.082>
- Le Gat, Y., Eisenbeis, P., 2000. Using maintenance records to forecast failures in water networks. Urban Water 2, 173–181.
[https://doi.org/10.1016/S1462-0758\(00\)00057-1](https://doi.org/10.1016/S1462-0758(00)00057-1)
- Lin, P., Yuan, X.-X., 2019. A two-time-scale point process model of water main breaks for infrastructure asset management. Water Res. 150, 296–309.
<https://doi.org/10.1016/j.watres.2018.11.066>
- Louppe, G., 2014. Understanding Random Forests: From Theory to Practice.
- Mailhot, A., Pelletier, G., Noël, J.-F., Villeneuve, J.-P., 2000. Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application. Water Resour. Res. 36, 3053–3062. <https://doi.org/10.1029/2000WR900185>

- Motiee, H., Ghasemnejad, S., 2019. Prediction of pipe failure rate in Tehran water distribution networks by applying regression models. *Water Supply* 19, 695–702. <https://doi.org/10.2166/ws.2018.137>
- Nishiyama, M., Fillion, Y., 2014. Forecasting breaks in cast iron water mains in the city of Kingston with an artificial neural network model. *Can. J. Civ. Eng.* 41, 918–923. <https://doi.org/10.1139/cjce-2014-0114>
- Nishiyama, M., Fillion, Y., 2013. Review of statistical water main break prediction models. *Can. J. Civ. Eng.* 40, 972–979. <https://doi.org/10.1139/cjce-2012-0424>
- Ofwat, 2020. Innovation in the water sector [WWW Document]. URL <https://www.ofwat.gov.uk/regulated-companies/innovation-in-the-water-sector/>
- Ponce Romero, J., Hallett, S., Jude, S., 2017. Leveraging Big Data Tools and Technologies: Addressing the Challenges of the Water Quality Sector. *Sustainability* 9, 2160. <https://doi.org/10.3390/su9122160>
- Rajani, B., Kleiner, Y., 2001. Comprehensive review of structural deterioration of water mains: Physically based models. *Urban Water* 3, 151–164. [https://doi.org/10.1016/S1462-0758\(01\)00032-2](https://doi.org/10.1016/S1462-0758(01)00032-2)
- Rajani, B., Kleiner, Y., Sadiq, R., 2006. Translation of pipe inspection results into condition ratings using the fuzzy synthetic evaluation technique. *J. Water Supply Res. Technol. - AQUA* 55, 11–24. <https://doi.org/10.2166/aqua.2005.069>
- Rajani, B., Kleiner, Y., Sink, J.-E., 2012. Exploration of the relationship between water main breaks and temperature covariates. *Urban Water J.* 9, 67–84. <https://doi.org/10.1080/1573062X.2011.630093>

- Rajani, B., Zhan, C., Kuraoka, S., 1996. Pipe soil interaction analysis of jointed water mains. *Can. Geotech. J.* 33, 393–404. <https://doi.org/10.1139/t96-061>
- Rezania, M., 2008. Evolutionary Polynomial Regression Based Constitutive Modelling and Incorporation in Finite Element Analysis. University of Exeter.
- 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. *Reliab. Eng. Syst. Saf.* 196, 106754. <https://doi.org/10.1016/j.ress.2019.106754>
- Robles-Velasco, A., Muñuzuri, J., Onieva, L., Rodríguez Palero, M., 2021. Trends and applications of machine learning in water supply networks management. *J. Ind. Eng. Manag.* 14, 45. <https://doi.org/10.3926/jiem.3280>
- Rodriguez, G., 2007. Survival Models, in: *Lecture Notes on Generalized Linear Models*. pp. 1–34.
- Røstum, J., 2000. Statistical Modelling of Pipe Failures in Water Networks. Norwegian University of Science and Technology.
- Sattar, A.M.A., Ertuğrul, Ö.F., Gharabaghi, B., McBean, E.A., Cao, J., 2019. Extreme learning machine model for water network management. *Neural Comput. Appl.* 31, 157–169. <https://doi.org/10.1007/s00521-017-2987-7>
- Scheidegger, A., Leitão, J.P., Scholten, L., 2015. Statistical failure models for water distribution pipes – A review from a unified perspective. *Water Res.* 83, 237–247. <https://doi.org/10.1016/j.watres.2015.06.027>
- Scheidegger, A., Scholten, L., Maurer, M., Reichert, P., 2013. Extension of pipe failure models to consider the absence of data from replaced pipes. *Water Res.* 47, 3696–3705. <https://doi.org/10.1016/j.watres.2013.04.017>

- Scholten, L., Scheidegger, A., Reichert, P., Maurer, M., 2013. Combining expert knowledge and local data for improved service life modeling of water supply networks. *Environ. Model. Softw.* 42, 1–16. <https://doi.org/10.1016/j.envsoft.2012.11.013>
- Shamir, U., Howard, C.D.D., 1979. An Analytic Approach to Scheduling Pipe Replacement. *J. Am. Water Works Assoc.* 71, 248–258. <https://doi.org/10.1002/j.1551-8833.1979.tb04345.x>
- Shirzad, A., Tabesh, M., Farmani, R., 2014. A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks. *KSCE J. Civ. Eng.* 18, 941–948. <https://doi.org/10.1007/s12205-014-0537-8>
- Snider, B., McBean, E.A., 2020. Watermain breaks and data: the intricate relationship between data availability and accuracy of predictions. *Urban Water J.* 17, 163–176. <https://doi.org/10.1080/1573062X.2020.1748664>
- Snider, B., McBean, E.A., 2019. Improving Urban Water Security through Pipe-Break Prediction Models: Machine Learning or Survival Analysis. *J. Environ. Eng.* 146, 04019129. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001657](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001657)
- Snider, B., McBean, E.A., 2018. Improving Time-to-Failure Predictions for Water Distribution Systems using Gradient Boosting Algorithm, in: 1st International WDSA / CCWI 2018 Joint Conference. Kingston, Ontario.
- Soguero-Ruiz, C., Mora-Jiménez, I., Mohedano-Munoz, M.A., Rubio-Sanchez, M., Miguel-Bohoyo, P. De, Sanchez, A., 2020. Visually guided classification trees for analyzing chronic patients. *BMC Bioinformatics* 21, 92. <https://doi.org/10.1186/s12859-020-3359-3>
- St. Clair, A.M., Sinha, S., 2012. State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models! *Urban Water J.* 9, 85–112. <https://doi.org/10.1080/1573062X.2011.644566>

- Tabesh, M., Soltani, J., Farmani, R., Savic, D., 2009. Assessing pipe failure rate and mechanical reliability of water distribution networks using data-driven modeling. *J. Hydroinformatics* 11, 1–17. <https://doi.org/10.2166/hydro.2009.008>
- Tang, K., Parsons, D.J., Jude, S., 2019. Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliab. Eng. Syst. Saf.* 186, 24–36. <https://doi.org/10.1016/j.res.2019.02.001>
- Wang, Y., Zayed, T., Moselhi, O., 2009. Prediction Models for Annual Break Rates of Water Mains. *J. Perform. Constr. Facil.* 23, 47–54. [https://doi.org/10.1061/\(ASCE\)0887-3828\(2009\)23:1\(47\)](https://doi.org/10.1061/(ASCE)0887-3828(2009)23:1(47))
- Watson, T.G., Christian, C.D., Mason, A.J., Smith, M.H., Meyer, R., 2004. Bayesian-based pipe failure model. *J. Hydroinformatics* 6, 259–264. <https://doi.org/10.2166/hydro.2004.0019>
- Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M., Zhao, X., 2018. A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* 82, 1027–1047. <https://doi.org/10.1016/j.rser.2017.09.108>
- Wilson, D., Fillion, Y., Moore, I., 2017. State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water J.* 14, 173–184. <https://doi.org/10.1080/1573062X.2015.1080848>
- Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., Tscheikner-Gratl, F., 2018. Pipe failure modelling for water distribution networks using boosted decision trees. *Struct. Infrastruct. Eng.* 14, 1402–1411. <https://doi.org/10.1080/15732479.2018.1443145>
- Wols, B.A., Vogelaar, A., Moerman, A., Raterman, B., 2019. Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. *Water Supply* 19, 404–416. <https://doi.org/10.2166/ws.2018.085>

- Xu, Q., Chen, Q., Li, W., 2011. Application of genetic programming to modeling pipe failures in water distribution systems. J. Hydroinformatics 13, 419–428. <https://doi.org/10.2166/hydro.2010.189>
- Yamijala, S., Guikema, S.D., Brumbelow, K., 2009. Statistical models for the analysis of water distribution system pipe break data. Reliab. Eng. Syst. Saf. 94, 282–293. <https://doi.org/10.1016/j.res.2008.03.011>

6 Predicting the Risk of Pipe Failure Using Gradient Boosted Decision Trees and Weighted Risk Analysis

Objective 4	<p>Chapter 6: Predicting the risk of pipe failure using gradient boosted decision trees and weighted risk analysis.</p>	<ul style="list-style-type: none">▪ Using the modelling approach identified in Objective 3 and the variables from Objective 2, predict pipe failures using a machine learning approach.▪ Develop suitable risk analysis to understand the risk of pipe failure.▪ Use GIS to visualise areas of the network at high risk in the event of pipe failure.
-------------	--	---

The following chapter fulfils Objective 4 and provides a methodology for predicting pipe failures and the risk of failure in drinking water networks, using the UK WDN case study. This model uses a gradient boosting tree model and weighted risk assessment. Alternative methods explored during the development of the model have been presented in Section D.2 and Section D.3. The following chapter has been taken directly, in line with the Creative Commons Attribution 4.0 international license agreement, from the following article published in npj Clean Water:

Barton, N.A., Hallett, S.H., Jude, S.R., Tran, T.H., 2022. Predicting the risk of pipe failure using gradient boosted decision trees and weighted risk analysis. npj Clean Water 5, 22. <https://doi.org/10.1038/s41545-022-00165-2>

Predicting the Risk of Pipe Failure Using Gradient Boosted Decision Trees and Weighted Risk Analysis

Neal Andrew Barton^a, Stephen Henry Hallett^{a*}, Simon Richard Jude^a and Trung Hieu Tran^b

^a School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

^b Centre for Competitive Creative Design (C4D), School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

**Corresponding author: Email: s.hallett@cranfield.ac.uk; Tel: +441234752978*

Abstract

Pipe failure prediction models are essential for informing proactive management decisions. This study aims to establish a reliable prediction model returning the probability of pipe failure using a gradient boosted tree model, and a specific segmentation and grouping of pipes on a 1 km grid that associates localised characteristics. The model is applied to an extensive UK network with approximately 40,000 km of pipeline and a 14-year failure history. The model was evaluated using the Receiver Operator Curve and Area Under the Curve (0.89), briers score (0.007) and Mathews Correlation Coefficient (0.27) for accuracy, indicating acceptable predictions. The acceptable accuracy was a result of the smaller spatial scale, short prediction interval, the absence of some essential variables, and inherent problems with pipe failure data. A weighted risk analysis

is used to identify the consequence of a pipe failure and provide a graphical representation of high-risk pipes for decision-makers. The weighted risk analysis provided an important step to understanding the consequences of the predicted failure. The model can be used directly in strategic planning, which sets long term key decisions regarding maintenance and potential replacement of pipes.

Key Words: Statistical analysis; clean water; pipe failure; infrastructure planning; machine learning

Highlights:

- Pipes are grouped by similar characteristics and based on a 1km grid.
- Gradient boosting trees can be used to model pipe failures with acceptable accuracy.
- Weighted risk analysis based on a combination of probability and consequence of failure.
- Risk map used for management decisions on network replacement, rehabilitation, or maintenance.

6.1 Introduction

Many Water Distribution Networks (WDN) are ageing and are in the closing phase of their design life, leading to pipe failures, leaks, and wasted water, causing environmental, economic, and societal impacts. With mounting pressure brought about by increased water demand and climate change impacts, leading to stress on water supply, and water regulators imposing heavy fines for not meeting performance objectives (Ofwat 2020), there is an urgency to reduce the effects of pipe failures through appropriate proactive management. Proactive management is the desired approach for managing WDNs, seeking to pre-empt issues and set levels of acceptable risk. Traditionally proactive management is achieved through prioritising the replacement or repair of pipes using simplistic rank likelihood models, expert judgment, and detailed network knowledge to

target critical areas of the network that have historically failed regularly (Snider and McBean, 2019). However, this simplistic approach is not suitable for managing WDNs with compound and complex risks. Proactive management requires understanding future pipe performance and assessing potential risk (Tang et al., 2019).

Statistical pipe failure models provide a means of supporting proactive management, predicting future performance through discerning failure patterns from historical data and their contributory causal factors (Chen et al., 2019). Shamir and Howard (1979) developed one of the first models on a small WDN of pipes, with a single-variate time-exponential model using pipe age to predict the number of failures per year per 1000 ft of pipe. Single-variate models are limited since multiple factors often operate concurrently to form complex mechanisms of failure that result in different modes of failure. These factors can be broadly categorised into pipe-intrinsic, environmental, and operational (Barton et al., 2019). Further progress was made using multivariate models, including static (pipe and soil) and dynamic time-related variables (weather) to predict the number of failures or failure rates (Kleiner and Rajani, 2000), that can be used to rank pipes against each other (Kakoudakis et al., 2018). Predicting the number of pipe failures at the asset level is mathematically problematic because incidents occur infrequently (Economou et al., 2012). Therefore, studies based on failure rate focus on grouping pipe failures by similar characteristics across a network, providing enough failures by grouped pipe length for statistical significance (Asnaashari et al., 2009; Giraldo-González and Rodríguez, 2020). However, grouping pipes at the network level makes assumptions that all pipes with similar characteristics share similar localised conditions (localised influences such as bedding conditions, traffic loading and local network jobs) and failure rates, which is rarely the case.

Deterministic models are reported to be too simplistic, being unable to handle randomness or left-truncated data normally found in pipe failure data sets which arise from missing failures due to limited short failure records (Snider and

McBean, 2020). Left-truncated data has the potential to mask pipes with high failure rates, and potentially reduce the accuracy of the final predictions. Furthermore, there is a need to predict failures at the pipe level to support effective management decision-making. Probability models such as survival analysis predict pipe failures at any phase in the life cycle (Wilson et al., 2017), consider the probability of random variables (St. Clair and Sinha, 2012), and use the likelihood function to accommodate left-truncated data (García et al., 2010). Two widely used survival analysis models include the Proportional Hazard (PH) methods of Cox PH and the Weibull PH, demonstrating good predictive accuracy compared to deterministic models (Kimutai et al., 2015; Snider and McBean, 2019). However, survival analysis is complex and only useful for long failure records, which many WDNs do not maintain (Chen et al., 2017). Other probability models include the use of the probability distribution, such as logistic regression. Both Motiee and Ghasemnejad (2019) and Yamijala et al. (2009) considered multiple models, including Poisson, linear, exponential, and logistic models developed to predict individual pipes. Both studies found logistic regression to provide the most useful results, since the probability of failure is often enough to inform management decisions and is more accurate than trying to predict the total number of failure at the pipe level, an approach where regression models have shown poor predictive (Chen et al., 2017; Motiee and Ghasemnejad, 2019; Tang et al., 2019; Yamijala et al., 2009). However, imbalanced data still presents a problem that should be carefully considered (Winkler et al., 2018). Kleiner and Rajani (2012) conclude that in general, due to inherent uncertainty and lack of data, analysing the behaviour of a single pipe is unfeasible. Therefore, it seems sensible to group pipes, but at a lower spatial level than the entire network. Few studies to date have attempted this, with Chen et al. (2019) being a notable exception, grouping at the census tract level (homogeneous areas based on number of people with respect to population characteristics, with an optimum size of 4,000 people (United States Census Bureau, 2021), leaving room for further exploration.

Machine learning models are becoming more widely used in pipe failure modelling, and preferred since mathematical processing steps are unnecessary (Winkler et al., 2018), with complex data summarised in a way that improves prediction accuracy, and with the tuning of interaction terms offering greater flexibility than with traditional models (Giraldo-González and Rodríguez, 2020). Machine learning methods are data-driven, and the methods most suited to modelling pipe failures are supervised since they are considered grey box approaches, allowing a degree of flexibility and suitability for structured data (Rudin, 2019). Commonly used supervised machine learning methods include artificial neural networks, evolutionary polynomial regression, and support vector machines. However, machine learning models are computationally expensive, especially when tuning several hyperparameters and have a limited scope for interpreting casual relationships between the response variable and covariates (Elith et al., 2008). Decision tree ensemble models overcome these limitations since they are more intuitive and transparent and can outperform other statistical methods. Studies have predominantly used Gradient Boosting Trees (GBT), which outperform other ensemble methods. Winkler et al. (2018), compared a Decision Tree, Random Forest, Adaboost and RUSboost model and found the RUSboost to have the highest accuracy (AUC of 0.93). Chen et al. (2019) compared a Gradient Boosting Model against a Generalised Linear Model, Generalised Additive Model, Random Forest, and Generalised Linear Mixed Models (GLMM). The authors concluded that the gradient boosting model performs well returning the lowest Brier scores between 0.558 and 0.808. Giraldo-González and Rodríguez (2020) compared a Support Vector Machine, Artificial Neural Network, Bayes, and Gradient Boosting model, and found the Gradient Boosting Model to perform best (AUC 0.998 for AC pipes and 0.990 for PVC pipes). These studies have typically focussed on five-year prediction intervals (Giraldo-González and Rodríguez, 2020; Winkler et al., 2018), or short monthly prediction intervals (Chen et al., 2019), yet some water network management decisions are performed annually, therefore it is sensible to understand the performance of GBT models on annual predictions.

For WDN managers, the concept of risk is important and yet often overlooked in pipe failure modelling (Suddle, 2009). Previous attempts to model risk include using the ranked ordered sorting of predictions based on the number of breaks (Aslani et al., 2021; Chen and Guikema, 2020) or the probability of failure (Giraldo-González and Rodríguez, 2020; Winkler et al., 2018). However, this approach is limited since water companies need to understand the risk of each potential failure as a combination of failure, either probability of failure or the number of failures, and the consequences. Christodoulou and Deligianni (2010) attempted to include a different risk level using proximity to buildings of public value and residential areas to prioritise repair and replacement work. Pietrucha-Urbanik and Tchórzewska-Cieślak (2018) proposed a framework for calculating the risk built on criteria grouping and weighting based on the potential financial losses arising. There are potentially numerous consequences of failures inherent to each network, yet common consequences include loss of water, potential disruption, reduction in water quality, reliability, direct costs (damage to property and infrastructure and pipe repair and replacement) and indirect costs (environmental and social). The risk of failure is complex, requiring several data sets from water companies and necessitates the difficult task of quantifying the consequences (Farewell et al., 2018). There is a gap in the literature in considering further developments in determining the risk of pipe failures.

Although many studies compare multiple models, it is difficult to ascertain which are superior, since WDN data is variable across networks and geographical regions, which cannot be captured by the model. Instead, model performance relies on data quality, availability, and the model development (Kerwin et al., 2020). Therefore, based on the gaps in the literature, this study aims to establish a reliable GBT prediction model for a UK WDN. The UK WDN has many of the typical issues that ageing infrastructure presents, which means most of the maintenance is performed reactively and wish to move towards proactively managing failures by predicting annual failures across the network. The WDN contains approximately 40,000 km of pipe covering some 27,476 km² of an urban

and rural environment, with a failure record history available over 14 years. The study will focus on the most commonly occurring pipe materials since the mechanism of failure has been established and can be accounted for by the variables used. The materials account for approximately 97 % of the UK WDN network, and include Iron, Steel and Ductile Iron (SDI), Asbestos Cement (AC), Polyvinyl Chloride (PVC) (collectively Unplasticised, Post Chlorinated and Molecular Orientated Polyvinyl Chloride) and Polyethylene (PE) (medium and high density). For shorter time intervals grouping pipes by similar characteristics is appropriate to yield more statistically accurate predictions, yet network wide groups are often unhelpful. This study uses a specific segmentation of pipes according to spatial characteristics and groups the segmented pipes by a 1 km interval. Using this 1 km interval is considered useful since it captures localised influences of weather and soil, removes the problem of grouping at larger spatial scale which often combines pipes with disparate failure rates, and presents smaller lengths of pipes and presents pipes with fewer failures, which is suitable for predicting the probability of failure. Previous studies have often limited failure models to predicting the probability of failure. Since the probability of failure alone is often not enough to support management decisions, this study builds on previous efforts by developing a practical approach to identify the risk of failure using weighted risk analysis.

6.2 Methodology

6.2.1 Decision Trees

The decision tree model is a machine learning method that is simple to implement, computationally efficient, and suitable for modelling complex relationships like those found in pipe failures (Winkler et al., 2018). A decision tree T partitions (or segments) the space of all explanatory variables into disjoint regions $R_1, R_2, \dots, R_j$ through recursive partitioning along the axis (known as axis parallel partitions) using a top-down greedy approach to identify regions within regions based on the Gini index, a measure of total variance across the classes.

The partitioning procedure continues until the stopping criterion is met, at which point the tree reaches the terminal node (the final space partitioned into non-overlapping regions). In this instance, the model describes the probability of failure via a Bernoulli distribution $P(x, y)$, where one indicates a certain failure and zero no failure. All probabilities returned are within this [one-zero] interval. A decision tree is formally described as (Hastie et al., 2009):

$$\hat{f}(x) = \sum_{j=1}^J c_R I\{(x_1, x_2, \dots, x_{14}) \in R_j\} \quad \text{Eq.6-1}$$

Where I is an indicator function, equal to 1 if the condition is true (failure) or 0 otherwise (non-failure). A constant c_R is applied to each partitioned region that determines the probability in that region.

Decision trees are relatively simple to interpret and visualise (Figure 6-1), can use multi-type variables, are not affected by variables on different scales, can accommodate missing variables and are insensitive to outliers. Yet decision trees model smooth functions poorly and can observe different partitions based on small changes to training data, introducing uncertainty, and resulting in poor predictions. It is therefore important to incorporate methods such as boosting to improve the predictions substantially.

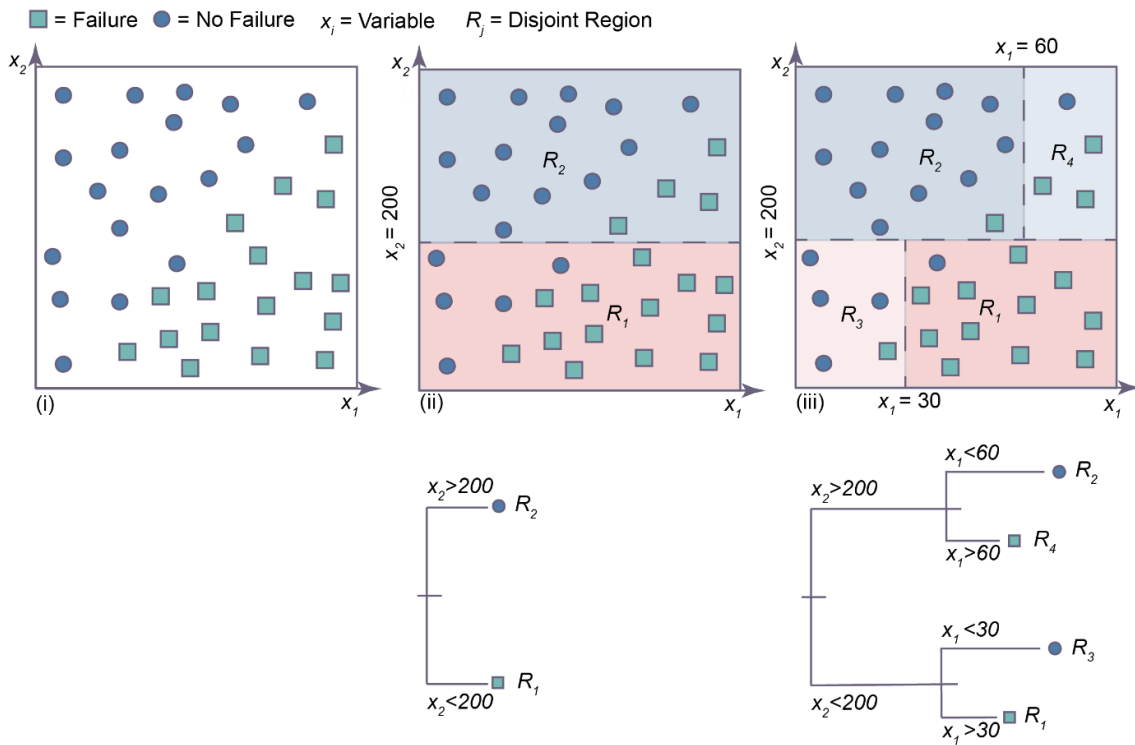


Figure 6-1 An example of data partitioning by a classification decision tree. (i) shows the two-dimensional data space. (ii) first condition for partitioning the data by variable x_2 at 200, where disjoint region R_1 is ≤ 200 and disjoint region R_2 > 200 . (iii) second condition for partitioning data by variable x_1 at both 30 and 60 to create two more disjoint region R_3 and R_4 (Taken from Barton et al. (2022)).

6.2.2 Gradient Boosting

Using an ensemble of trees proves beneficial since the model learns slower and reduces overfitting variance and bias (Hastie et al., 2009). One such ensemble model is gradient boosting, a form of functional gradient descent, which describes a forward stage-wise process fitting multiple trees iteratively to the training data, aiming to minimise the loss function in the existing collection of trees by adding, at each step, another tree that best reduces the loss function. The loss function is a measure of how well the model coefficients fit the data, and in this study, the negative gradient of the deviance is used, which for classification models is the residual of the response minus the fitted probability mean, where $r = y - \hat{y}$. The process is described by building a function $\hat{f}_B(x)$ which is the sum of the tree

ensemble. The first tree is fitted with boosting iteration m_1 to the training data and the response y , maximally reducing the loss function, from which the residuals are determined as $r_1 = y_1 - \hat{y}_1$. Subsequent trees are fitted in the same way, but the following trees are updated based on the residuals of the previous trees such that $r_i = r_{n-1} - \hat{r}_{n-1}$. Overfitting is avoided using regularization applied as a shrinkage penalty factor of $0 < \lambda < 1$ to scale the contribution of the tree. Regularisation through shrinkage offers a robust alternative to traditional variable selection methods (Elith et al., 2008). Regularization necessitates mutually optimising the number of trees, the tree complexity, and the learning rate. Another advantage of regularization is that several covariates can be included in the model, and if they have a limited effect on the response, their contribution will simply be down weighted. This is easier than adding and removing variables to build a parsimonious model (Elith et al., 2008). The shrunken tree is then added to the function: $\hat{f}_B(x) \leftarrow \hat{f}_B(x) + \lambda T(x; \gamma)$ where x is the multivariate argument characterised by a set of parameters γ . The following trees with boosting iteration $b_2, b_3, \dots, B$ are trained using the training data, and the residuals of each tree iteratively. Each tree is shrunken and successively added to the function, and the residuals updated so that $r_i \leftarrow r_i - T(x; \gamma_b)$. The final regression gradient boosting model is depicted in Figure 6-2 and can be expressed as (Hastie et al., 2009):

$$\hat{f}_B(x) = \sum_{b=1}^B T(x; \gamma_b) \quad \text{Eq.6-2}$$

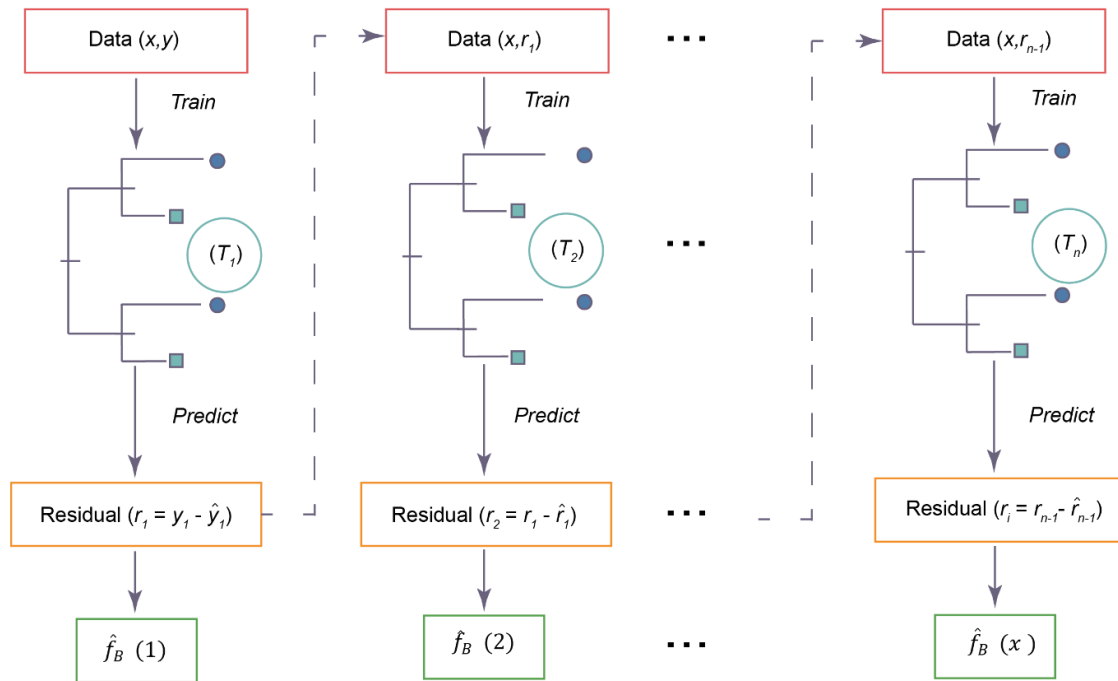


Figure 6-2 Gradient Boosting Tree ensemble model (Taken from Barton et al. (2022))

The gradient boosting model has many hyperparameters that control the execution of the learning. A sequential grid search across the different hyperparameters was undertaken to optimise the performance and to yield the best model. Each hyperparameter was tuned using an appropriate range, and the number of trees used in the boosting ensemble increased until the results no longer improved (Elith et al., 2008). Five-fold cross-validation was used to balance the computational complexity of the model and its accuracy. The technique for K -fold cross-validation randomly divides the training data into K equal subsamples, where a single subsample is retained for testing and the remaining subsamples are used for training (Elith et al., 2008). The process is repeated K times, using each subsample as the testing subsample at least once. Cross-validation calculates multiple estimates of 'out of sample error', returning the smallest to minimise overfitting (Hastie et al., 2009). R version 3.6.2 was used to develop the models (R Core Team, 2018). The 'gbm' package version 2.8.1

(Greenwell et al., 2019), and 'caret' package version 6.0 (Kuhn et al., 2020) were both used from the CRAN repository.

6.2.3 Weighted Risk Analysis

There are several ways to determine risk (Pietrucha-Urbanik and Tchórzewska-Cieślak, 2018). In this study, risk R is a combination of the failure probability P^f and the sum of the consequences $\sum_i C_i^f$, i.e., water loss (pipe diameter) and flooding damage (proximity to nearest property). The weighting is associated with the importance of each consequence. Since there is often more than one consequence, the consequences are summed and weighted by importance (Suttle, 2009). The final calculation can be expressed as $R = P^f \times \sum_i C_i^f$. Table 6-1 shows the consequences, their weights and associated severity scores. The consequence score was determined by using four categories. The diameters were categorised according to diameter bands, and an assumed increase in water loss for larger diameters. The potential damage to property was determined in conjunction with expert knowledge, estimating that a pipe failure will likely cause more damage to closer properties. Given that approximately 71% of the network has a pipe diameter of <166 mm, catastrophic events and large volumes of water loss are unlikely; therefore, properties within 10m are at highest risk.

Table 6-1 Consequences of failure, their weights and associated severity scores.

Consequence	Weighting	Units	Consequence Score
Diameter – volume of water lost*	0.4	< 166 mm	0.25
		166 - < 321 mm	0.5
		321 - < 626 mm	0.75
		≥ 626 mm	1
Distance to property – chance of damage and disruption	0.6	≥ 10	0.25
		> 5 to ≤ 10 m	0.5
		> 2m to ≤ 5 m	0.75
		≤ 2 m	1

Weighted risk analysis uses the probability of failure, the diameter size, and the pipe’s proximity to the nearest property since these were the only data available. The units determined for the consequence score were determined through dialogue with risk managers at UK water utility companies. The distance from the pipe to the nearest property was calculated using OS OpenMap buildings (Ordnance Survey, 2021) and the GIS package ArcGIS Pro (ESRI, 2021) to calculate the shortest planar distance between the pipe and the nearest property. The outcome of the weighted risk analysis is presented using Natural Jenks to arrange the data into three categories of risk, low, medium, and high. Natural Jenks is a clustering method that seeks to minimise the average deviation in each class based on natural groupings inherent in the data. Natural Jenks are advantageous since they identify real classes within the data and provide more meaningful visualisations (North, 2009).

6.2.4 Data Curation

The methods were applied to a UK WDN, operating over an area of approximately 27,476 km² and supplying approximately 4.3 million people with drinking water. The UK WDN dataset includes failure records collected between 2005 and 2018, with information on pipe location, length, material type, age, diameter and water source, and failure location and time. The pipe failures collected on site were often geographically displaced from the pipe failure event. Therefore, all pipe failures were relocated to the nearest pipe to ensure no data was lost. Each pipe failure was firstly relocated within 3 m (a distance accounting for GPS error), and if no match was made, then the process was repeated sequentially up to 1 km until a pipe with equivalent characteristics of diameter and material type was found. Table 6-2 shows a summary of the WDN data.

Table 6-2 WDN data collected between 2005 and 2018.

	Iron	AC	SDI	PVC	PE
Length of Pipe ¹	11,382	7,090	1,824	5,854	12,274
Failure rate (failures/km/year)	0.27	0.17	0.05	0.17	0.03

1 Length of network operational by the end of December 2018.

Temperature and Soil Moisture Deficit (SMD) data were attained from the Met Office Rainfall and Evaporation Calculation System (MORECS 40 x 40 km grid) in a weekly summarised format. The total number of days air frost data were sourced from the Met Office summary data sets (Met Office, 2018), downloaded in a monthly summarised format. The soil data collected from the national soil map-related Natural Perils Directory and LandIS soil data and maps from Cranfield University (Hallett et al., 2017) are presented as 1:250,000 maps based on field data collected between 1939 and 1987. Using ArcGIS Pro, the pipe network data was segmented by the underlying soil characteristics, and the associated soil data attributed to the pipe segment. The MORECS and summary weather data were added to the dataset based on the 40 × 40 km MORECS grid value that was associated with the pipe using R software (version 4.0.0), and the pipe diameter and age were placed into categorised bands. The final covariates shown in Table 6-3 were selected based on available data and on those factors known to correlate with pipe failures, as discovered in complementary studies undertaken by the authors (Barton et al., 2020, 2019).

Table 6-3 Feature vector of variables selected for the pipe failure model, including description and data type.

Variable	Description	Type
Material	Pipe materials – Iron, AC, SDI, PVC, PE.	Categories
Age	Partitioned into age since installation categories of 20 years.	Categories
Diameter	Four diameter band ordered categories comprising: < 166 mm, 166 - < 321 mm, 321 - < 626 mm and ≥ 626 mm.	Categorical
Length	Total pipe length	Numeric (km)
Number of Previous Failures	Total number of previous failures for each group of pipes	Numeric (total failures)
Water Source	The originating source of water is divided into three categories, groundwater (GW), surface water (SW) and mixed water (MW).	Categorical
Temperature	The total number of weeks per year is below 5 °C, and the total number of weeks per year with a temperature greater than 20 °C.	Numeric (weeks)
Soil Moisture Deficit	The total number of weeks with SMD greater than 120 mm. SMD - the difference between soil moisture content and soil moisture saturation.	Numeric (weeks)
Days Air Frost	Total annual days air frost.	Numeric (days)
Shrink Swell	Six categories of potential soil volumetric expansion < 3 %, 3-5 %, 5-12 %, 12-15 % and 12-15* % (when soils are drained to 2 m) ¹ .	Categorical
Hydrology of Soil Type	Describes the dominant hydrology type of the soil in eight categories ² .	Categorical
Soil Corrosivity	Categorical data on the corrosivity of the soil to Iron. Six ordered categories ³ .	Categorical
Clay Content	Dominant clay vulnerability class between 1 – 9 - indicates the level of movement and potential ground subsidence ⁴ .	Categorical

¹ See Table D-1 of Appendix D

² See Table D-2 of Appendix D

³ See Table D-3 of Appendix D

⁴ See Table D-4 of Appendix D

The pipes are segmented and grouped based on similar characteristics, including material, diameter band, age band, soil characteristics, expressed in a 1 km grid to capture localised conditions, and remove the failure rate disparities shared when grouping pipes across a network. Each weather variable is summarised into extreme weather conditions (maximum and minimum values) and joined to the dataset. The final dataset contains 80,107 cohorts, with an average length of 433 m, a minimum length of 2m and a maximum length of 11,995 m. The data is imbalanced, with one or more failures representing only 0.1%. Since the purpose is to predict the probability of pipe failure, which is typically enough information for decision makers, the number of failures are substituted for either failure or no failure. Some studies have separated material types into distinct data sets for modelling, since the mechanisms of failure are often unique to each material type. Yet here the data is used in a global model that included all materials, since several studies have suggested global models are the most suitable approach (Jafar et al., 2010; Robles-Velasco et al., 2020; Winkler et al., 2018) for three main reasons; (1) many variables are unavailable that are specifically unique to each material therefore most of the variables influence all materials, (2) the most unique aspect of the materials is the seasonal difference in failure rate, which, due to the annual predictions, is not included here, and (3) some materials such as SDI do not have enough pipe failures for good model convergence, yet through a global model this problem is removed by learning from a greater number of failures.

The data was partitioned into 70 % training and 30 % testing, a common approach for this type of study (Giraldo-González and Rodríguez, 2020; Kimutai et al., 2015; Motiee and Ghasemnejad, 2019), where large training data sets have shown improved model performance (Snider and McBean, 2020). Randomly partitioning over the time frame is also useful since partitioning by year may introduce bias in the model in particularly extreme years (e.g. the hottest year on record) (Kabir et al., 2016; Yamijala et al., 2009). Stratified random sampling was used during the partition to ensure a representative sample of each material was

included in both the training and testing data set, such that: $N = \sum_{i=1}^k N_i$, where k are the number of strata (in this case the five materials) and N_i the number of sampling units in the i^{th} strata.

6.3 Results and Discussion

6.3.1 Receiver Operator Curve and Area Under the Curve

The receiver operator curve (ROC) is used to visualise how the model performs independently of the decision threshold, providing a useful tool for visualising how well the classifier avoids false classifications (Saito and Rehmsmeier, 2015). The ROC plot shows a trade-off between the True Positive Rate (TPR) or sensitivity, the fraction of observations that are correctly classified, calculated as:

$$TPR = \frac{TP}{TP + FN} \quad \text{Eq.6-3}$$

where TP is True Positive and FN False Negative, and False Positive Rates (FPR) or specificity, the fraction of observations that are incorrectly classified, calculated as:

$$FPR = \frac{FP}{FP + TN} \quad \text{Eq.6-4}$$

The passing of two lines corresponding to a 100 % TPR and a 0 % FPR = 1 (TPR versus 1 – FPR) is considered a perfect discriminatory ability. This is graphically represented by the ROC curve passing the top left-hand corner of the plot. The passing of the curve through the diagonal $y = x$ represents a model that is no better than a random guess (Fawcett, 2006). The Area Under the Curve (AUC) is an aggregated measure of performance for all classification thresholds and represents the measure of separability by describing the capability of the predictions in distinguishing between the classes. An AUC measure is returned between zero and one, with zero representing a perfectly inaccurate test and one a perfect test. In general, an AUC of 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and > 0.9 is outstanding (Hosmer and Lemeshow,

2000). Figure 6-3 shows the ROC curve for the test data set close to the top left-hand corner and an AUC value of 0.89, suggesting the model has an excellent discriminative ability to distinguish between the classes, and the TPR and FPR appear robust enough to predict failures on the unseen test data.

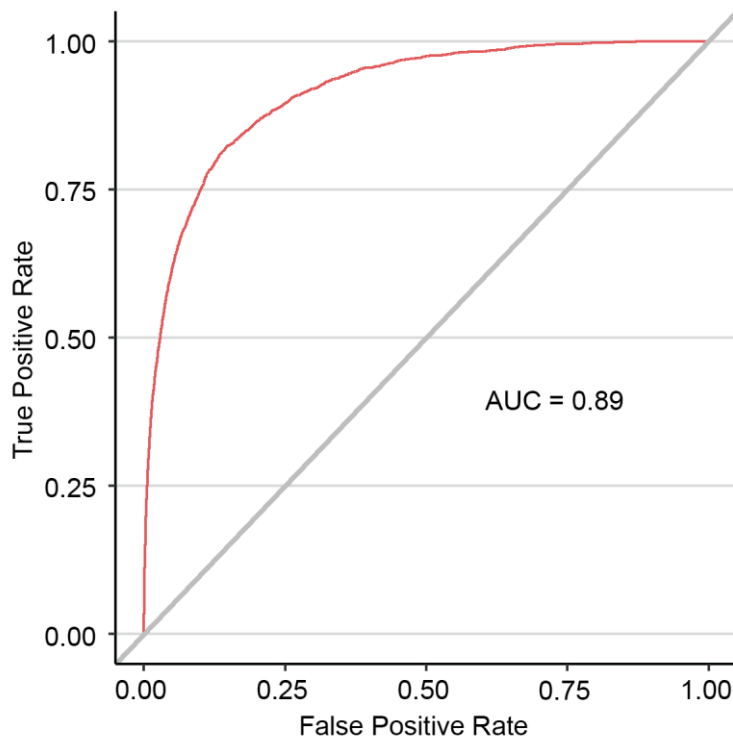


Figure 6-3 Test data accuracy, Receiver Operator Curve (ROC) curve with Area Under the Curve (AUC) measure of performance for all classification thresholds. The red line is the ROC curve, and the grey line represents the diagonal $y = x$ and a point where the curve is random.

The calibration curve provides a means of observing how close the predictions are to those observed. Since the outcome in this model is the probability of failure between 0 and 1, it is appropriate to use a binning method. Binning is advantageous since it averages the probability of failure for each bin which provides a useful graphical representation of how well the model is calibrated. The mean probability is then compared to the frequency of observed failures in each bin. In this case, a fixed-width binning approach is used, where the data is

partitioned into ten bins known as decile analysis, and approach used in similar studies (Kumar et al., 2018). A reliability curve provides a means of visualising this comparison, whereby perfectly calibrated probabilities would lie on a diagonal line through the middle of the plot. The briers score is a useful measure of accuracy for probabilistic predictions and is equivalent to the mean squared error whereby the cost function minimises to zero for a perfect model and maximises to 1 for a model with no accuracy (Chen et al., 2019). The Brier's Score (BS) is calculated as:

$$BS = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \quad \text{Eq.6-5}$$

where N is the total number of observations, P_i is the prediction probability and O_i is the equal to the event outcome failure or no failure. Figure 6-4 Test data accuracy, calibration curve with Brier's score. The red line is the calibration curve; the grey line represents a perfect fit. shows the calibration plot for the model and suggests the model is well calibrated for the lower and upper deciles since most bins fit the diagonal. The upper middle deciles do not fit the diagonal where the calibration curve is below or above the diagonal, suggesting the predictions have a lower probability than those seen in the data The briers score of 0.007 is low, suggesting accurate predictions overall.

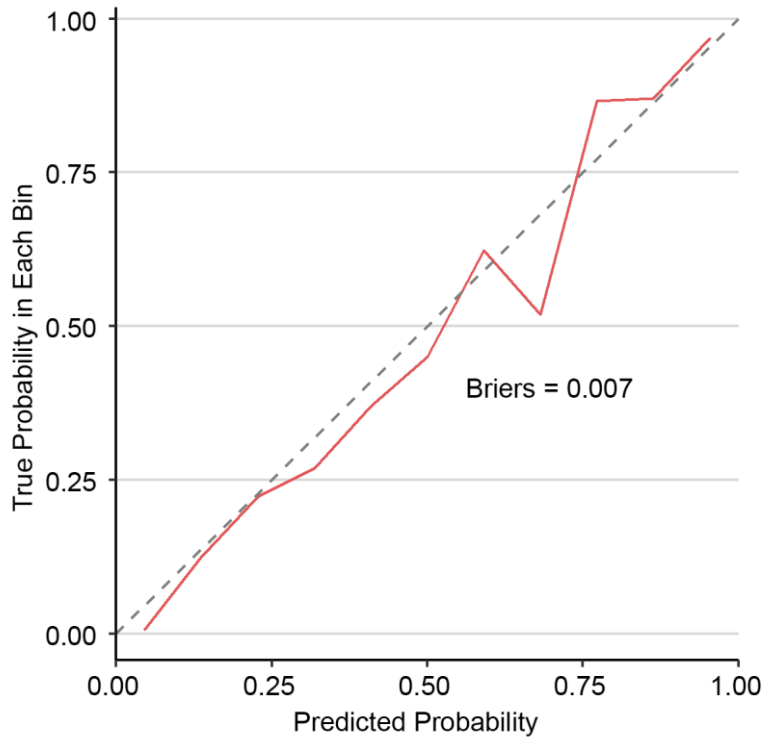


Figure 6-4 Test data accuracy, calibration curve with Brier's score. The red line is the calibration curve; the grey line represents a perfect fit.

6.3.2 Confusion Matrix and Accuracy

The confusion matrix describes the frequency of classification outcomes by explicitly defining the number of True Positives (TP, or Precision), True Negatives (TN), False Positives (FP), and False Negatives (FN). The decision to convert a predicted probability into a class label is determined by an optimal probability threshold such that the value of the response:

$$y_i = \begin{cases} \text{no failure if } P_i \leq \text{threshold} \\ \text{failure if } P_i > \text{threshold} \end{cases} \quad \text{Eq.6-6}$$

The default probability threshold within the model is 0.5 (Robles-Velasco et al., 2020). By this definition, there remains a practical need to optimise the probability threshold specifically to the behaviour of pipe failures within the imbalanced test

data. An optimal probability threshold typically strikes a balance between sensitivity and specificity. However, there is a trade-off between TPR and FPR when altering the threshold, where increasing or decreasing the TPR typically results in the same for the FPR and vice versa. Probability threshold optimisation is an important step in the decision-making process and is specific to each problem. In the case of pipe replacement, expert judgement should be used by reasoning that water companies would seek to avoid unnecessarily replacing pipes that may have a longevity of several decades more, resulting in wasted maintenance effort and cost. Furthermore, only 0.5–1% of the network is typically replaced each year due to budget constraints (Bruaset and Sægrov, 2018). It is therefore important to only identify pipes with the highest probability of failure. Considering this, the optimal threshold is set to reduce the FNs (i.e., pipes predicted to fail when they have not). This reduces the number of TPs predicted as discussed above but targets those pipes most likely to fail.

A factorial experimental design was used, whereby the threshold was iterated from 0.01 through to 0.99, observing each threshold to reveal the point where the highest accuracy meets the lowest FN value. The Matthews Correlation Coefficient (MCC) was used to measure accuracy and is useful for imbalanced data since it accounts for the difference in class size and only returns a high accuracy score if all four confusion matrix categories are accurately represented. For this reason, Chicco (2017) argues that it is the correct measure for imbalanced data sets. The MCC describes the prediction accuracy as worst value = -1 and best value = +1 and is calculated as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad \text{Eq.6-7}$$

Table 6-4 Table of thresholds from training data, including True Positive, False Positive, False Negative, True Negative, Matthews Correlation Coefficient, percentage of pipe identified with a true failure, and the percentage of true failures predicted correctly. Only a few rows are presented for succinctness. Bold row text shows the final threshold chosen.

Threshold	True Positive (TP %)	False Positive (FP %)	False Negative (FN %)	True Negative (TN %)	MCC	% of pipe in WDN with failures	% of total failures in the WDN captured
0.10	37.19	62.81	1.29	98.71	0.26	9.33	23.37
0.11	34.81	65.18	1.11	98.89	0.27	8.55	21.74
0.12	32.80	67.20	0.95	99.05	0.27	7.83	20.20
0.13	30.39	69.61	0.83	99.17	0.26	7.22	18.78
0.14	27.99	72.01	0.73	99.27	0.26	6.70	17.31

Table 6-4 shows a small range of the thresholds for brevity. The optimal threshold in this instance has been identified firstly with the highest MCC accuracy and then the lowest FN. The MCC of 0.27 suggests the model is better than a random fit, but a low MCC value also represents a high percentage of false positives (i.e., values incorrectly identified as non-failure). The balanced accuracy for this model is 0.65. In practical terms, the results are helpful for water companies to target areas for further investigation and potential replacement since they focus on those pipes having the highest probability of failure, yet there are still incorrect predictions that could lead to the potential replacement of pipes unnecessarily.

The model predicts 20.20% of all failures occurring in the WDN, found in 7.83% of the WDN pipe network. The results show that approximately 32.80% of the observed pipe failures were correctly predicted as failures, whilst approximately 67.20% of the observed pipe failures were falsely predicted as no failure. If desired, water companies could choose an alternative threshold, one that eliminates FN predictions, however, the number of TP predictions will also reduce.

6.3.3 Relative Variable Importance

The relative variable influence shows the empirical improvement I_t^2 accounted for by variable interval x_j , averaged across all boosted trees as presented:

$$\hat{f}_j^2 = \sum_{\text{splits on } x_j} I_t^2 \quad \text{Eq.6-8}$$

The variable influence helps understand which variables contribute more when predicting pipe failures. For GBT models, this is the summation of predictor influence accumulated over all the classifiers. Figure 6-5 shows the results, suggesting similar findings compared to existing literature. The most important variables are the number of previous failures and pipe length, both a proxy for pipe performance and deterioration. It is worth reiterating that both variables represent the grouped pipe and do not consider individual pipe history. Soil Moisture Deficit (SMD) is the most important weather variable being linked with shrinkage of clay soils and subsequent ground movement in AC pipe failures. Conversely, clay soils and soils shrink–swell potential, both representing ground movement, show lower influence.

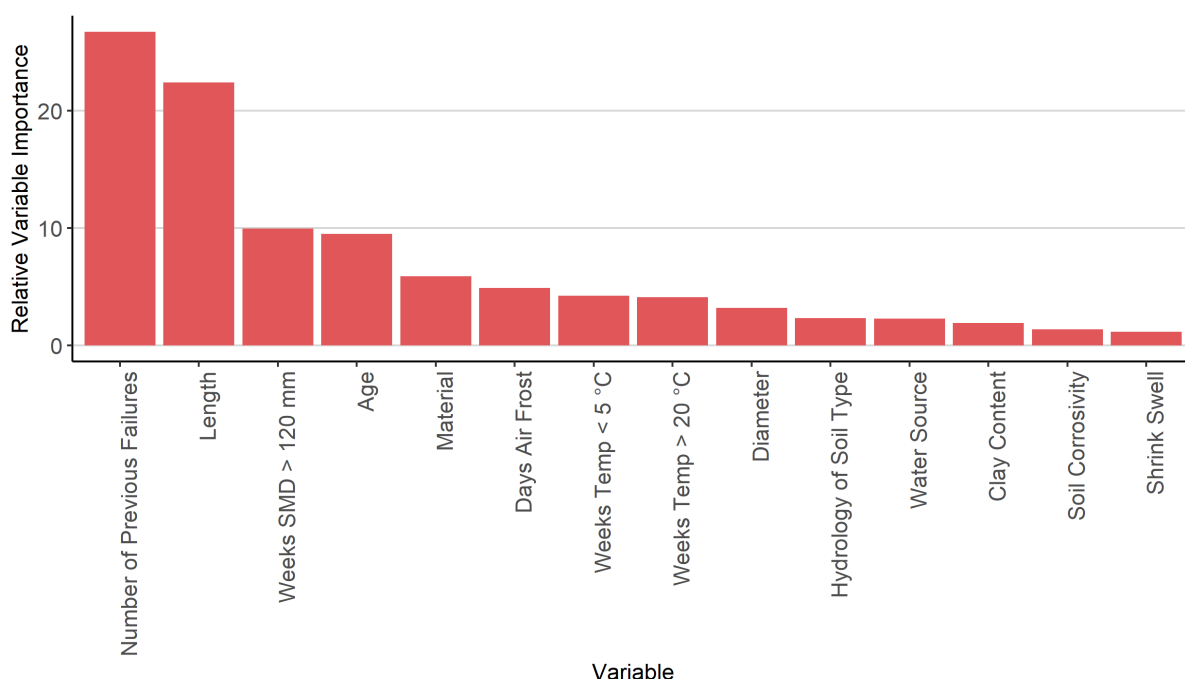


Figure 6-5 Relative Importance of the variables used in the global model.

Pipe diameter, and material are less important factors in this network than as reported in comparable studies (Giraldo-González and Rodríguez, 2020; Kleiner and Rajani, 2012; Konstantinou and Stoianov, 2020; Winkler et al., 2018). The relative variable influence of days air frost and temperature is not as high as expected, given their correlation with high pipe failure frequency in iron pipes and the large percentage of iron pipes in the WDN. It is likely to be a result of over summarising the data to facilitate the annual prediction interval. A shorter prediction interval (week, or month) for networkwide groups of pipes is necessary to capture inter-annual variation accurately, but short prediction intervals in the authors' experience can result in low predictive accuracy. The overall relative variable influence of soil (shrink well, soil corrosivity, Hydrology of Soil Type) is low. From literature and an engineering perspective, soil corrosion is strongly related to the deterioration of metal pipes and their ability to withstand internal and external forces (Tang et al., 2019). It is possible that many pipes in this network may have been rehabilitated and protected against corrosion; however, this information was unavailable at the time of this study. Water source is the only

operational variable, and shows low importance compared to many other variables. The most important water source is surface water, resulting in lower temperatures during the winter due to its exposure to weather. This causes higher failure rates in metal pipes, yet compared to other variables, the importance is low. Other variables are imaginable such as installation details like bedding and backfill material, surrounding environments providing evidence on loading such as traffic loading and construction works, operational data such as pipe pressure and transients, water quality and spatial failure characteristics. These are not investigated here but will likely result in performance gains.

6.3.4 Risk Mapping

For the mapping to be effective from an asset management standpoint, the results of the weighted risk analysis should be able to separate out low, medium, and high failures. The number of high failures is expected to be small for two reasons, (1) pipes rarely fail more than once and (2) utilities are only able to allocate investments to those at the greatest risk due to budget limitations and are therefore only interested in the top 1–2% of pipes. The outcome of the weighted risk analysis is presented in Figure 6-6, representing a small section of the WDN for clarity. Natural Jenks arranges the risk level into three categories, low $[0; \leq 0.02]$, medium $[>0.02; \leq 0.06]$ and high $[>0.06; \leq 0.92]$. In this scenario, the length of pipe in the high-risk category is 13.9 km of the 300.7 km or 4.6% of the pipe network present in Figure 6-6, useful percentage of the network to target for management decisions. The choropleth risk map approach is an important means of visualising individual pipes or clusters of pipes with the highest risk in the WDN, evidenced in Figure 6-6. Figure 6-6 also highlights how many pipes in this section of the network have a low risk, which is to be expected since many pipes have a low probability of failure and have small diameters, potentially causing less damage if they fail.

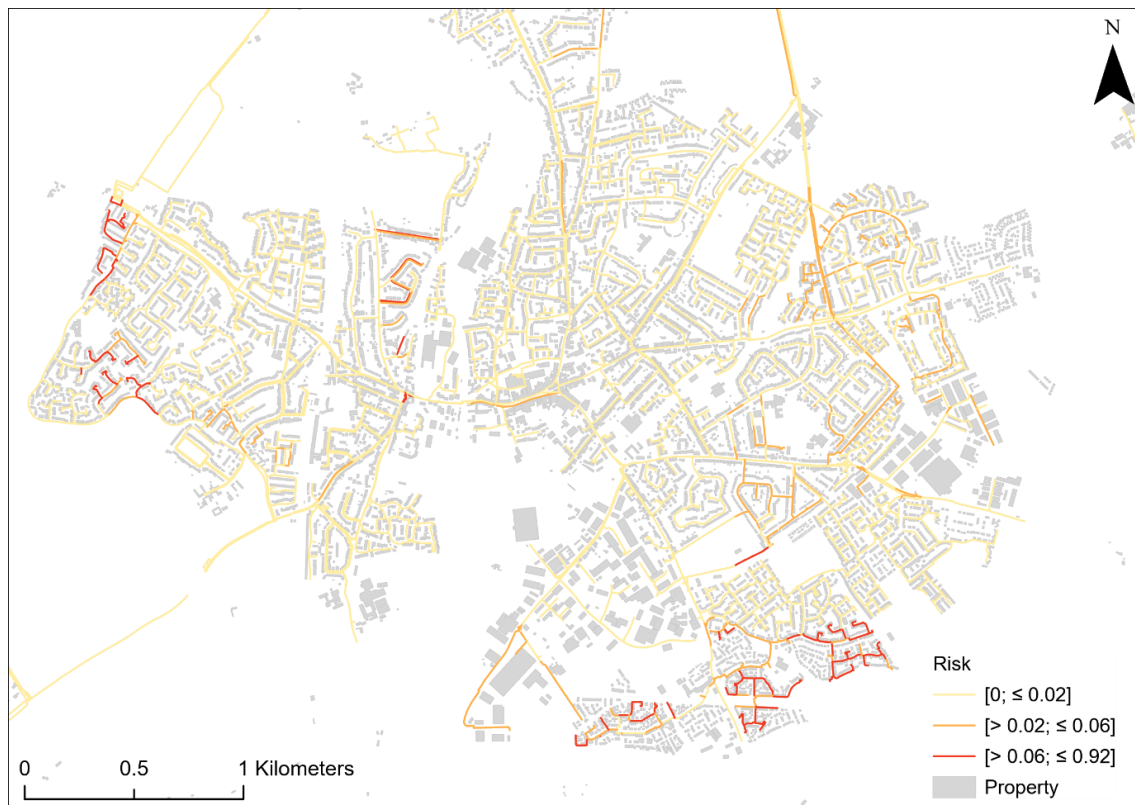


Figure 6-6 Choropleth weighted risk map categorised using Natural Jenks. Risk is calculated as a measure of the probability of pipe failure and the consequence of damage to nearest property and water lost based on pipe diameter. The map represents a small section of the entire UK WDN.

6.3.5 Practical Considerations

Creating groups of pipes was an important step given the low frequency of failures in the UK WDN dataset. Grouping pipes in this way assumes that all pipes in the group share similar failure rates, which is not the case, and thus the approach adopted here presents a suitable solution to this limitation. Grouping pipes on a lower spatial scale can capture localised influences on pipe performance, that can often be obfuscated when generalising over the whole network. However, the approach used may not be as useful for rural areas where fewer pipes are present, where smaller scales may be more appropriate (e.g., 1:100,000 is a smaller scale than 1:100). Further investigation into grouping scales is merited. Optimisation the threshold is challenging and inevitably leads

to inappropriately classified failures on either side of the threshold. Optimising is even more difficult with imbalanced data sets since conventional classification methods are built to assume that all classes are equal. An alternative approach was applied in this study, which used MCC accuracy and FN to set a threshold, reducing the potential for wasting budgets replacing pipes that will not fail. In the process, the number of TPs was reduced to 32.80% of the observed pipe failures, whilst the number of FPs was 67.20% of the observed pipe failures, which may not present a good argument to professionals. Despite this, the results can be used directly in strategic planning, which sets long-term key decisions regarding maintenance and potential replacement of pipes. Predicting the probability of failure is an essential response since it enables the identification and prioritisation of risk across the network. This methodology could also be used to provide longer-term predictions to support the development of Asset Management Plan, which cover a five-year period of regulated investment.

Categorising the pipes based on a weighted risk analysis and visually presenting them using Natural Jenks offers a useful method for prioritising pipes based on the consequence of their failure and is an easily assessed cartographic presentation. It extends the probability of failure into a more useful measure of risk, providing more information for decision makers. The use of distance to property in this study is a simple approach to determine flooding. To provide a realistic determination of flooding, an understanding of key geographical features for overland flow routing is required (Ozcelik and Gorokhovich, 2020). The list of consequences was limited in this study and could be extended when such data is available. There are potentially numerous consequences of failure inherent to each network, yet common consequences include loss of water, potential disruption, reduction in water quality, reliability, direct costs (damage to property and infrastructure and pipe repair and replacement) and indirect costs (environmental and social) (Kakoudakis et al., 2018). In this study, the risk estimates were concluded by expert knowledge, and any contextual mismatch between weightings could potentially skew the outcomes. Therefore, the

weightings should be considered carefully by network professionals. At an engineering level, the risk mapping can be further used to determine areas of the network leading to a high probability of failure, which can be used to take constructive pre-emptive actions towards extending the life of future pipe construction (Fitchett et al., 2020).

The economic benefits of this model will manifest when performing proactive maintenance, potentially averting associated risks that may arise from damaging properties and infrastructure. It is anticipated that the modelling approach proposed will enhance decision-making at a local level, facilitated through numerical outputs which report on the serviceability of the WDN and help meet regulatory performance targets avoiding heavy fines. Operationally, the approach will help with highlighting short pipe segments for repair and replacement though graphical outputs, these are practical lengths of pipes for operational teams that typically do not replace kms of pipe at any given time (De Oliveira et al., 2011a). This approach shows similar performance to comparable GBT studies (Giraldo-González and Rodríguez, 2020; Winkler et al., 2018) but is beneficial since the method provides reliable predictions on a shorter annual time frame. The method here is also computationally easier to develop than other more complex machine learning methods such as neural networks and Bayesian Neural Networks.

The predictions rely on the quality of the data, and several challenges were presented during the cleaning and processing, most notably the location of the pipe failures, many of which were geographically displaced, and some by a considerable distance yet was necessary to retain all the failures in the dataset. These were snapped to the nearest pipe with similar characteristics, yet it is conceivable that some were incorrectly placed despite the protocols established for the snapping process. Further limitations to the study include limited data, where pressure data or other operational data may have proved useful, the advantage of which may consist of increased model accuracy and interpretability. Over-summarised local conditions can also affect the model accuracy, and in this study, the local soil conditions were presented from a soil map at 1:250,000 scale.

Likewise, the weather variables were highly summarised to an annual scale from a 40 × 40 km grid source. Inevitably these limitations will affect the model, which can potentially hinder effective decision making. There are several challenges faced when modelling pipe failures, from uncertainties in data collection and management to specific data processing solutions. There is a need to understand these holistically, and from the view of current practice for a more in-depth perspective of current challenges in practice that may hinder useful data gathering. In addition, future research aimed at understanding how practitioners understand pipe failure models, the limitations, and opportunities is beneficial, since there is often a discord between the capabilities of modelling and user expectations. This further research may help to improve pipe failure models by encouraging enhancements in the pipe failure model process that promotes quality data capture.

6.3.6 Conclusion

This study considered the prediction of pipe failures using a GBT model and establishing the risk based on weighted risk analysis to prioritise pipes for proactive management. A 1 km spatial scale was included in this model when grouping the pipes, which aimed to capture localised conditions and remove the failure rate disparities shared when grouping pipes across a network. This spatial scale, together with a short prediction interval, the absence of some essential variables, and additional inherent problems with pipe failure data sets, has ultimately resulted in acceptable accuracy. However, in practical terms, when used in conjunction with expert knowledge, the results provide a useful approximation of potential failures and a better understanding of the current WDN to help plan rehabilitation and replacement efforts. Improving model accuracy may be achieved by increasing the prediction interval to five-year asset management plan, potentially accumulating more failures per pipe group from which to predict. Yet this may not be as useful to water companies where management decisions are typically annual. Furthermore, understanding the issues faced with data collection and quality from current practice may help to

encourage data quantity and quality, and could potentially provide marked improvements in the final predictions.

Further suggested research includes exploring different pipe grouping variations, collecting more data on the consequences of failure to enhance the weighted risk analysis and, expanding on this idea, understanding the data quantity and quality issues from current practice, and exploring feature engineering techniques to derive more valuable data sets that may improve model accuracy.

6.4 Author Contributions

Neal Barton: Conceptualisation, Methodology, Software, Testing, Formal Analysis, Investigation, Resources, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualisation and Project administration. Stephen Hallett: Writing – Review & Editing, Supervision, Funding Acquisition. Simon Jude: Review & Editing, Supervision, Funding Acquisition. Trung Hieu Tran: Testing, Formal Analysis and Review & Editing.

6.5 References

- Asnaashari, A., McBean, E.A., Shahrour, I., Gharabaghi, B., 2009. Prediction of watermain failure frequencies using multiple and Poisson regression. *Water Supply* 9, 9–19. <https://doi.org/10.2166/ws.2009.020>
- Barton, N.A., Farewell, T.S., Hallett, S.H., 2020. Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *npj Clean Water* 3, 31. <https://doi.org/10.1038/s41545-020-0077-3>
- Barton, N.A., Farewell, T.S., Hallett, S.H., Acland, T.F., 2019. Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Res.* 164, 114926. <https://doi.org/10.1016/j.watres.2019.114926>

- Barton, N.A., Hallett, S.H., Jude, S.R., Tran, T.H., 2022. An evolution of statistical pipe failure models for drinking water networks: a targeted review. *Water Supply*. <https://doi.org/10.2166/ws.2022.019>
- Berardi, L., Giustolisi, O., Kapelan, Z., Savic, D.A., 2008. Development of pipe deterioration models for water distribution systems using EPR. *J. Hydroinformatics* 10, 113–126. <https://doi.org/10.2166/hydro.2008.012>
- Bruaset, S., Sægrov, S., 2018. An analysis of the potential impact of climate change on the structural reliability of drinking water pipes in cold climate regions. *Water (Switzerland)* 10. <https://doi.org/10.3390/w10040411>
- Chen, T.Y.-J., Beekman, J.A., David Guikema, S., Shashaani, S., 2019. Statistical Modeling in Absence of System Specific Data: Exploratory Empirical Analysis for Prediction of Water Main Breaks. *J. Infrastruct. Syst.* 25, 04019009. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000482](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000482)
- Chen, T.Y., Beekman, J.A., Guikema, S.D., 2017. Drinking Water Distribution Systems Asset Management: Statistical Modelling of Pipe Breaks, in: *Pipelines 2017*. American Society of Civil Engineers, Reston, VA, pp. 173–186. <https://doi.org/10.1061/9780784480885.017>
- Chen, T.Y.J., Guikema, S.D., 2020. Prediction of water main failures with the spatial clustering of breaks. *Reliab. Eng. Syst. Saf.* 203, 107108. <https://doi.org/10.1016/j.res.2020.107108>
- Chicco, D., 2017. Ten quick tips for machine learning in computational biology. *BioData Min.* 10, 35. <https://doi.org/10.1186/s13040-017-0155-3>
- Christodoulou, S., Deligianni, A., 2010. A Neurofuzzy Decision Framework for the Management of Water Distribution Networks. *Water Resour. Manag.* 24, 139–156. <https://doi.org/10.1007/s11269-009-9441-2>
- De Oliveira, D.P., Garrett, J.H., Soibelman, L., 2011. A density-based spatial clustering approach for defining local indicators of drinking water

- distribution pipe breakage. *Adv. Eng. Informatics* 25, 380–389. <https://doi.org/10.1016/j.aei.2010.09.001>
- Economou, T., Kapelan, Z., Bailey, T.C., 2012. On the prediction of underground water pipe failures: zero inflation and pipe-specific effects. *J. Hydroinformatics* 14, 872–883. <https://doi.org/10.2166/hydro.2012.144>
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- ESRI, 2021. ArcGIS Pro [WWW Document]. URL <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview> (accessed 2.16.21).
- Farewell, T.S., Jude, S., Pritchard, O., 2018. How the impacts of burst water mains are influenced by soil sand content. *Nat. Hazards Earth Syst. Sci.* 18, 2951–2968. <https://doi.org/10.5194/nhess-18-2951-2018>
- Fawcett, T., 2006. An introduction to ROC analysis. *Pattern Recognit. Lett.* 27, 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Fitchett, J.C., Karadimitriou, K., West, Z., Hughes, D.M., 2020. Machine Learning for Pipe Condition Assessments. *J. Am. Water Works Assoc.* 112, 50–55. <https://doi.org/10.1002/awwa.1501>
- García, A.C., Solano, H., Gamiz, M.L., Debón, A., 2010. Evaluation of the Reliability of a Water Supply Network from Right-Censored and Left-Truncated Break Data. *Water Resour. Manag.* 24, 2917–2935. <https://doi.org/10.1007/s11269-010-9587-y>
- Giraldo-González, M.M., Rodríguez, J.P., 2020. Comparison of Statistical and Machine Learning Models for Pipe Failure Modeling in Water Distribution Networks. *Water* 12, 1153. <https://doi.org/10.3390/w12041153>
- Greenwell, B., Boehmke, B., Cunningham, J., 2019. Package ‘gbm’ - Generalized Boosted Regression Models. CRAN Repos.

- Hallett, S.H., Sakrabani, R., Keay, C.A., Hannam, J.A., 2017. Developments in land information systems: examples demonstrating land resource management capabilities and options. *Soil Use Manag.* 33, 514–529. <https://doi.org/10.1111/sum.12380>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning*, Springer Series in Statistics, Springer Series in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-84858-7>
- Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*. John Wiley & Sons, Inc., Hoboken, NJ, USA. <https://doi.org/10.1002/0471722146>
- Jafar, R., Shahrour, I., Juran, I., 2010. Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Math. Comput. Model.* 51, 1170–1180. <https://doi.org/10.1016/j.mcm.2009.12.033>
- Kabir, G., Tesfamariam, S., Loepky, J., Sadiq, R., 2016. Predicting water main failures: A Bayesian model updating approach. *Knowledge-Based Syst.* 110, 144–156. <https://doi.org/10.1016/j.knosys.2016.07.024>
- Kakoudakis, K., Farmani, R., Butler, D., 2018. Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. *J. Hydroinformatics* 20, 1191–1200. <https://doi.org/10.2166/hydro.2018.152>
- Kerwin, S., Garcia de Soto, B., Adey, B., Sampatakaki, K., Heller, H., 2020. Combining recorded failures and expert opinion in the development of ANN pipe failure prediction models. *Sustain. Resilient Infrastruct.* 1–23. <https://doi.org/10.1080/23789689.2020.1787033>
- Kimutai, E., Betrie, G., Brander, R., Sadiq, R., Tesfamariam, S., 2015. Comparison of Statistical Models for Predicting Pipe Failures: Illustrative Example with the City of Calgary Water Main Failure. *J. Pipeline Syst. Eng. Pract.* 6, 04015005. [https://doi.org/10.1061/\(ASCE\)PS.1949-1204.0000196](https://doi.org/10.1061/(ASCE)PS.1949-1204.0000196)

- Kleiner, Y., Rajani, B., 2012. Comparison of four models to rank failure likelihood of individual pipes. *J. Hydroinformatics* 14, 659–681. <https://doi.org/10.2166/hydro.2011.029>
- Kleiner, Y., Rajani, B., 2000. Considering time-dependent factors in the statistical prediction of water main breaks, in: American Water Works Association Infrastructure Conference (AWWA 2000). Seattle, USA, pp. 1–12.
- Konstantinou, C., Stoianov, I., 2020. A comparative study of statistical and machine learning methods to infer causes of pipe breaks in water supply networks. *Urban Water J.* 17, 534–548. <https://doi.org/10.1080/1573062X.2020.1800758>
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Ziem, A., Scrucca, L., Hunt, T., 2020. Package ‘caret’ - Classification and Regression Training.
- Kumar, A., Rizvi, S.A.A., Brooks, B., Vanderveld, R.A., Wilson, K.H., Kenney, C., Edelstein, S., Finch, A., Maxwell, A., Zuckerbraun, J., Ghani, R., 2018. Using Machine Learning to Assess the Risk of and Prevent Water Main Breaks, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, New York, NY, USA, pp. 472–480. <https://doi.org/10.1145/3219819.3219835>
- Met Office, 2018. UK regional climate summaries 1981 - 2018 [WWW Document]. Met Off. URL <https://www.metoffice.gov.uk/research/climate/maps-and-data/regional-climates/index>
- Motiee, H., Ghasemnejad, S., 2019. Prediction of pipe failure rate in Tehran water distribution networks by applying regression models. *Water Supply* 19, 695–702. <https://doi.org/10.2166/ws.2018.137>
- North, M.A., 2009. A Method for Implementing a Statistically Significant Number of Data Classes in the Jenks Algorithm, in: 2009 Sixth International

- Conference on Fuzzy Systems and Knowledge Discovery. IEEE, pp. 35–38. <https://doi.org/10.1109/FSKD.2009.319>
- Ofwat, 2020. PN 22/18: Thames Water’s failure to tackle leakage results in £65m package for customers [WWW Document]. URL <https://www.ofwat.gov.uk/pn-22-18-thames-waters-failure-tackle-leakage-results-65m-package-customers/> (accessed 1.5.21).
- Ordnance Survey, 2021. OS OpenMap - Local [WWW Document]. URL <https://www.ordnancesurvey.co.uk/business-government/products/open-map-local> (accessed 4.10.21).
- Ozcelik, C., Gorokhovich, Y., 2020. An overland flood model for geographical information systems. Water (Switzerland) 12. <https://doi.org/10.3390/W12092397>
- Pietrucha-Urbanik, K., Tchórzewska-Cieślak, B., 2018. Approaches to Failure Risk Analysis of the Water Distribution Network with Regard to the Safety of Consumers. Water 10, 1679. <https://doi.org/10.3390/w10111679>
- R Core Team, 2018. R: The R Project for Statistical Computing [WWW Document]. URL <https://www.r-project.org/> (accessed 2.6.19).
- Ridgeway, G., 2019. Generalized Boosted Models: A guide to the gbm package.
- 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. Reliab. Eng. Syst. Saf. 196, 106754. <https://doi.org/10.1016/j.ress.2019.106754>
- Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Saito, T., Rehmsmeier, M., 2015. The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced

- Datasets. PLoS One 10, e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Shamir, U., Howard, C.D.D., 1979. An Analytic Approach to Scheduling Pipe Replacement. J. Am. Water Works Assoc. 71, 248–258. <https://doi.org/10.1002/j.1551-8833.1979.tb04345.x>
- Snider, B., McBean, E.A., 2020. Watermain breaks and data: the intricate relationship between data availability and accuracy of predictions. Urban Water J. 17, 163–176. <https://doi.org/10.1080/1573062X.2020.1748664>
- Snider, B., McBean, E.A., 2019. Improving Urban Water Security through Pipe-Break Prediction Models: Machine Learning or Survival Analysis. J. Environ. Eng. 146, 04019129. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001657](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001657)
- St. Clair, A.M., Sinha, S., 2012. State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models! Urban Water J. 9, 85–112. <https://doi.org/10.1080/1573062X.2011.644566>
- Suddle, S., 2009. The weighted risk analysis. Saf. Sci. 47, 668–679. <https://doi.org/10.1016/j.ssci.2008.09.005>
- Tang, K., Parsons, D.J., Jude, S., 2019. Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. Reliab. Eng. Syst. Saf. 186, 24–36. <https://doi.org/10.1016/j.res.2019.02.001>
- United States Census Bureau, 2021. Glossary [WWW Document]. URL [https://www.census.gov/programs-surveys/geography/about/glossary.html#:~:text=Census tracts generally have a, optimum size of 4%2C000 people. \(accessed 1.28.22\).](https://www.census.gov/programs-surveys/geography/about/glossary.html#:~:text=Census tracts generally have a, optimum size of 4%2C000 people. (accessed 1.28.22).)

- Wilson, D., Filion, Y., Moore, I., 2017. State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water J.* 14, 173–184. <https://doi.org/10.1080/1573062X.2015.1080848>
- Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., Tscheikner-Gratl, F., 2018. Pipe failure modelling for water distribution networks using boosted decision trees. *Struct. Infrastruct. Eng.* 14, 1402–1411. <https://doi.org/10.1080/15732479.2018.1443145>
- Yamijala, S., Guikema, S.D., Brumbelow, K., 2009. Statistical models for the analysis of water distribution system pipe break data. *Reliab. Eng. Syst. Saf.* 94, 282–293. <https://doi.org/10.1016/j.res.2008.03.011>

7 The Challenges of Predicting Pipe Failures in Clean Water Networks: A View from Current Practice

Objective 5	<div><p>Chapter 7: The challenges of predicting pipe failures in clean water networks: a view from current practice</p><ul style="list-style-type: none">▪ Using the results and understanding from Objectives 3 and 4. Detail the challenges and limitations of pipe failure data and modelling.▪ Understand the issues faced in industry by collecting qualitative data from semi-structured interviews with industry professionals.</div>
-------------	--

The following chapter fulfils Objective 5 by discussing the limitations of pipe failure modelling, including semi-structured interviews with industry professionals. The chapter is taken directly, without further edits, from the following article published in the International Water Associations Journal, Water Supply, in line with the Creative Commons Attribution 4.0 international license agreement:

Barton, N.A., Hallett, S.H., Jude, S.R., 2021. The challenges of predicting pipe failures in clean water networks: a view from current practice. Water Supply. <https://doi.org/10.2166/ws.2021.255>

The Challenges of Predicting Pipe Failures in Clean Water Networks: A View from Current Practice

Neal Andrew Barton^a, Stephen Henry Hallett^{a*} and Simon Richard Jude^a

^a School of Water, Energy and Environment, Cranfield University, Bedfordshire, MK43 0AL, United Kingdom.

**Corresponding author: Email: s.hallett@cranfield.ac.uk; Tel: +441234752978*

Abstract

Pipe failure models can aid proactive management decisions and help target pipes in need of preventative repair or replacement. Yet, there are several uncertainties and challenges that arise when developing models, resulting in discord between failure predictions and those observed in the field. This paper aims to raise awareness of the main challenges, uncertainties, and potential advances discussed in key themes, supported by a series of semi-structured interviews undertaken with water professionals. The main discussion topics include data management, data limitations, pre-processing difficulties, model scalability and future opportunities and challenges. Improving data quality and quantity is key in improving pipe failure models. Technological advances in the collection of continuous real-time data from ubiquitous smart networks offer opportunities to improve data collection, whilst machine learning and data analytics methods offer a chance to improve future predictions. In some instances, technological approaches may provide better solutions to tackling short-term proactive management. Yet, there remains an opportunity for pipe

failure models to provide valuable insights for long-term rehabilitation and replacement planning.

Key Words: Clean water; Data analytics, Infrastructure planning, Smart networks, Water supply.

Highlights:

- A variety of pipe failure model challenges and uncertainties are discussed.
- Industry professional interviews provide deeper understanding of the issues.
- Data management, quality, pre-processing, and scalability influence modelling.
- Future opportunities include the benefits of real-time data, data analytics and machine learning.

7.1 Introduction

Pipe failures result in billions of litres lost from water networks each day, which wastes water, causes damage to infrastructure, and interrupts continuous service. Considerable effort is being concentrated on minimising water loss and improving services through performance commitments aimed to reduce pipe failures (Ramirez et al., 2020; Robles-Velasco et al., 2020). There has been a recent focus on innovation, partly driven by advances in technology and through regulator incentives (Ofwat, 2020c) as a way of catalysing change in the industry. In consequence, water companies are moving towards data analytics and statistical models (referred to hereafter as pipe failure models) to provide insights for proactive management decisions. Pipe failure models represent a distinct field of data analytics, and although commencing with Shamir and Howard (1979), they are still considered innovative today with the emergence and application of machine learning approaches and big data.

Pipe failure models provide a means to understand future network performance, discerning patterns from historical data, describing the response as either the probability of failure (probability), time-to-failure (survival analysis), pipe failure rates (regression) or whether or not a failure will occur (classification) (Economou et al., 2012). Many studies report useful insights that inform proactive (also known as preventative) management decisions. Nonetheless, uncertainties are an integral part of prediction modelling, and underlying many pipe failure studies are issues surrounding data quality and quantity and model development. As such, there remains a degree of scepticism of pipe failure models (Kleiner and Rajani, 2012; Konstantinou and Stoianov, 2020) and the discord between failure predictions and those observed in the field, potentially hindering effective decision-making.

Water networks have complex failure histories, and studies have often reported poor data quality (Kabir et al., 2016; Mailhot et al., 2000; Tang et al., 2019). In particular, the timing and location of pipe failures (Wu and Liu, 2017), the limited failure history and absence of decommissioned pipe data, and limited variables leading to a lack of understanding about the various factors influencing pipe failures (Scheidegger et al., 2015). Model scalability is perhaps the most complex challenge in pipe failure assessment, a theme widely discussed in the literature due to the gravitas of its inherent effect on predictive accuracy (Kleiner and Rajani, 2012). Due to these challenges, pre-processing steps are often required, including relocating of pipe failures, handling of missing and inaccurate data, and scaling the model appropriately by employing pipe grouping methods at the network or census level (Chen et al., 2019), k-means clustering (Kakoudakis et al., 2018), or re-cutting pipes to improve model accuracy (Winkler et al., 2018). Understanding these challenges is vital for both managers and professionals involved in developing pipe failure models, yet whilst literature has acknowledged some difficulties, to our knowledge, there are few studies explicitly discussing these challenges holistically. Furthermore, it is warranted that these issues should be further explored from the perspective of industry professionals, which

to our knowledge has not been documented. This perspective offers a more in-depth understanding of the challenges.

While general study limitations and uncertainties are reported in the literature, it is important to consider the issues faced, from data collection and management to specific data processing solutions, all of which affect the predicted outcome. Thus, this study discusses a holistic view of the various challenges, providing further understanding and insight from current practices in water companies through a series of semi-structured interviews. Accordingly, the study is discussed in five key themes: (i) challenges of data management, (ii) limitations of pipe failure data, (iii) complexities of pre-processing, (iv) understanding model scalability, and (v) future opportunities and challenges.

7.2 Methodology

Water professionals from UK water companies were interviewed using a semi-structured approach to reveal views on the selected themes. The aim was to identify key issues and challenges faced to complement and support the topic discussions formed from the literature and an understanding of pipe failure models. The semi-structured interviews were based on the practical guidelines detailed in Adams (2015). The interview guide was pre-defined (see Appendix E.2) based on previous knowledge of the subject area (Barton et al., 2020; 2019) and the desire to explore topics based on these findings. In this respect, the methodological approach adopted is similar to deductive qualitative content analysis, whereby the structure of the analysis was based on previous knowledge and the interview guide designed to understand and characterise extant themes and situations. In addition to the themes the participants were also asked questions on the model developed in Chapter six, with an aim to gauge the operational practicality of the methodology. The authors use manifest analysis, whereby quotes of participants are used to provide further detailed understanding on a theme of interest (Bengtsson, 2016). The semi-structured nature allowed participants, by design, the freedom to diverge and elaborate beyond a certain

point, revealing further information of relevance (Adams, 2015). Consent was obtained from all participants, and anonymity provided. A pilot semi-structured interview based on the questions was completed on two water industry professionals before commencement on the primary cohort of participants. The semi-structured interviews were undertaken between January 2021 and April 2021 and involved eight participants drawn from different UK companies. Each interview was restricted to approximately one hour to reduce participant and interviewer fatigue (participant details are provided in Table 7-1). The interviews were recorded and transcribed so relevant information for each of the themes could be included.

Table 7-1 Participants of the semi-structured interview.

ID	Area of expertise in clean water network management	Sector	Interview duration
P1	Data scientist	Water utility	1:07:23
P2	Network modeller	Water utility	59:58
P3	Leakage analyst	Water utility	52:09
P4	Network modeller	Water utility	55:53
P5	Asset manager	Consultant	57:12
P6	Leakage analyst	Consultant	58:41
P7	Data scientist	Water utility	1:01:02
P8	Network manager	Water Utility	45:14

The number of participants in semi-structured interviews is often discussed, but it is common for studies to have between one and 30 participants (Bengtsson, 2016). Galvin (2015) reviewed several engineering articles and found the average number of participants was between eight and 17, but further found several examples where the number of participants was lower, as low as two in some cases. For homogenous groups where all participants had a familiarity within the study area, Guest et al. (2006) found a lower number of participants was sufficient. Noting this, the number of participants should be determined

based on conclusive answers to the research questions with sufficient confidence, to the point of information saturation, which occurs when no new information is being observed (Bengtsson, 2016; Fusch and Ness, 2015; Ritchie et al., 2003). Many professional practitioners were asked to participate in this study. The participants selected had no previous involvement in the project and were recruited through 'snowball' methods. Eight identified themselves as topic experts. The mix of data scientists, asset managers and leakage analysts constitute enough participants to collect the desired information, mainly due to the distinct topics, the participants' deep expertise providing a rich quality and quantity of information, and the length and design of the interviews.

7.3 Results and Discussion

The results are structured based on the key themes outlined in the introduction, divided further by areas of interest derived from the interviews. The themes and illustrative quotes identify the challenges and uncertainty seen in pipe failure models from the perspective of professionals. Figure 7-1 shows a systematic map of the themes and sub-themes discussed hereafter.

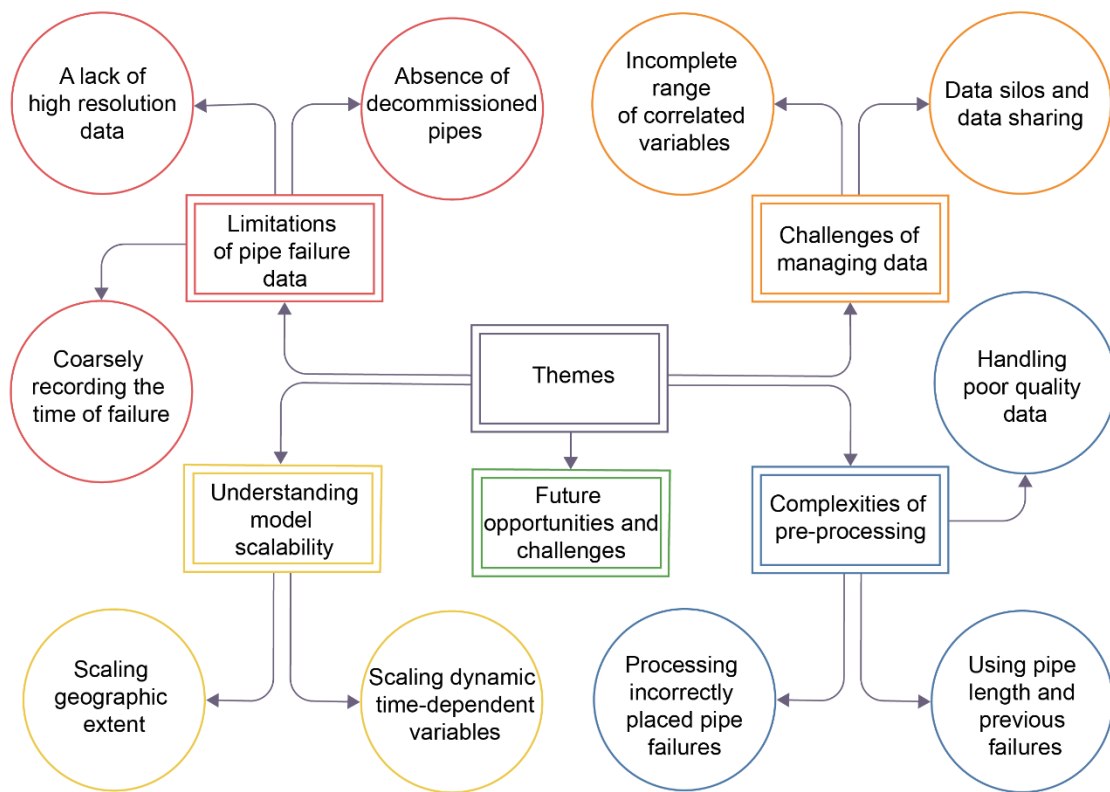


Figure 7-1 Systematic map of the themes and sub-themes explored in this study.

7.4 Challenges of Managing Data

Many Water Distribution Networks (WDN) consist of an antiquated asset base nearing the end of their engineering life span (Snider and McBean, 2019). Over the past century, upgrades, repairs, and industry regulation and management changes have led to a complex non-linear network comprised of discontinuous pipe materials, requiring careful documentation of asset status information. Over time, data collected on the constantly evolving WDNs have grown organically with the needs of utility companies, from the original paper asset records to more widely captured contemporary data using advanced real-time sensors and digital storage. Whilst current data is rigorously checked and is generally reliable, much of the information on the historical network construction remains limited, often in analogue form, or of poor quality and is often highly variable in terms of data accuracy, coverage, scale, and format, which significantly limits its quality, accessibility, and usability (Likhari et al., 2017). All participants acknowledged

historical data quality as an issue, with one suggesting *“The actual quality of the [historical] data will be limited.”* [P4]. The unreliable nature of the data creates two apparent problems. Firstly, the lack of historical failure data results in imbalanced data that is problematic to model. Due to the lack of historical data, attempts have been made at pooling data from different water companies to increase the size of data sets, such as the UKWIRs National Mains Failure Database (NMFD) (UKWIR, 2020). Yet much effort would need to be expended to clean, process, and harmonise the disparate databases, often collected through different formats and processes unique to each water company. Previous attempts to model using this data have shown poor results (Tang et al., 2019). Renaud et al. (2007) tried a similar approach by combining data from water companies across France to overcome the scarce data issue and found no marked improvement. Secondly, there is a question of geodetic reliability and accuracy of pipe locations recorded before using Geographical Information Systems (GIS) and Global Navigation Satellite System (GNSS). Participants highlighted the issues with historical pipe locations, with one highlighting the problem with finding pipes when they failed, *“Trying to locate this old pipe, even with ground penetrating radar, we had no chance of finding it at all.”* [P6]. Where pipes are accurately located, there is speculation over the level of accuracy in the information, such as the recorded material or the pipe diameter with one participant suggesting *“What material will be laid won’t be accurate, water companies will do their best, but it won’t all be right.”* [P4]. Similar observations were made on the network data, which is constantly evolving and updating making it difficult to maintain the new asset data and pipe replacements, meaning failure models based on older network data will not be applied to the most recent network and may be identifying pipes already replaced or rehabilitated, *“The GIS system should have accurate existing data on an asset, but there are multiple versions across departments where work is constantly being done.”* [P2]. However, it was suggested by the respondents that considerable progress has been made in this respect, and water companies are making concerted efforts to improve network data, mainly aided by advances in GIS.

The accuracy of historical data is essential to modern analytical approaches. While problems persist, it is evident most water companies have directed resources into ensuring the best quality possible. Yet the quality of historical data will remain an issue: *“The problem is the [historical] data can't change, and it's very difficult to change it retrospectively. So, you're only really talking about improving the way things get done in the future and not improving past data.”* [P5]. Most importantly, all participants had commenced concerted efforts to capture future data more accurately, acknowledging the importance of data accuracy for gaining valuable insights.

In the UK, the water sector was privatised and in the process government regulators were introduced, which led to significant improvements in data collection and quality (Ofwat, 2006) but in turn presented new challenges. The introduction of Asset Management Plan (AMP) cycles was created as a means of permit regulatory price reviews and performance assessment. Regulatory changes between AMP cycles and the need to deliver business value and efficiency whilst reporting to the regulatory authority have resulted in associated variations in data collection and subsequent gaps in data, inhibiting its collective usability (Tang, 2018). When discussed with participants, several suggested data was collected foremost to meet regulatory requirements, as noted by several participants:

“Regulatory periods can drive data collection, or whatever the business sees as cost saving or improving business efficiency.” [P5].

“It will be down to the individual water company what data they collect, but as long as they can report to Ofwat the total number of failures for assets and the length of the assets that is kind of the limit that is imposed.” [P1].

“Some managers will say we will no longer collect that data since it's no longer reportable, without understanding the value of it for the business in going forward, which lost several years of information.” [P6].

Future data management methods should prioritise continuous and consistent data collection across regulatory periods, although one participant noted “*Getting new information relies heavily on budget.*” [P8], which presents another problem. Nonetheless, ongoing data collection will provide prolonged periods of data collection, which is essential. Some respondents noted that many operatives were not made aware of the importance of data - in which case, there is no incentive to keep collecting it. Communication is the key to ensuring operatives understand the use of the data. Continuous data collection is also achievable by adopting ubiquitous smart sensor technologies, such as multi-sensor platforms or fibre optics (e.g., Distributed Acoustic Sensors (DAS)). Such approaches will result in complete data capture without gaps. Such data characteristics often require data standardisation and processing to be performed more efficiently, so expertise with big data technologies should be considered (Sun and Scanlon, 2019).

7.4.1 An Incomplete Range of Correlated Variables

Predicting pipe failures requires several variables correlated with the various modes and mechanisms of failure (for further detail see Barton et al. (2019; 2020)). Collecting variables is a complex task, and pipe failure models are ordinarily developed based on limited data because variables are not routinely collected, are unnecessary for regulatory reporting requirements or immediate business needs, or are considered too costly to acquire due to budget restrictions (Ramirez et al., 2020). As one participant noted “*In a situation where water companies are squeezed for budget, things that look like a ‘nice to have’ get squeezed out.*” [P1]. Where data is available, it is often framed for a particular use elsewhere in the business, meaning additional time spent on data cleaning and transformation. The involuntary omission of important variables results in variable bias, preventing the estimator from converging correctly, causing an inaccurate representation of pipe failures (Konstantinou and Stoianov, 2020; Tang et al., 2019). If pipe failure models are to aid decision makers effectively,

water companies must prioritise data collection, ensuring enough data related to pipe failures is readily available and collected in useful formats.

7.4.2 Data Silos and Data Sharing

Data silos still exist widely in water companies due to evolving operations, creating many specialist areas and multiple management layers, leading to barriers inhibiting data sharing. Open data policies and incentives have been shown to lead to profound benefits, primarily using data lakes that store and transfer data more efficiently. However, this approach can create additional complexity, cost, and latency, worsened by increased data volumes (Sun and Scanlon, 2019). Participants acknowledged that data silos exist as a means of working, one participant mentioned *“Different teams have different datasets, where the rest of the teams don’t have access to it or don’t even know it exists.”* [P8]. However, several suggested that data lakes widely exist, offering broader data sharing opportunities. In this case, participants highlighted difficulties accessing the data, especially when using systems that require specialist support, with one remarking *“We have a specialist team who look after the data...we can request the information at any time, but it’s difficult because the team is always really busy.”* [P1]. Where specialist teams exist to help with data extraction and transformation, it is evident that difficulties persist due to a lack of professional informatics staff. In these instances, data were requested, kept, and updated locally since this proved easier, but in doing so subsequently creating multiple versions of the data. Issues were also identified with communicating the data lake and its intended use, with one participant noting *“It’s misunderstood and not accessed by everyone who could benefit from it”* [P4]. Further problems regarding the misuse of data were noted as a problem:

“I think that you’d have to make sure that there was a good description of what all the data sets were and what they mean. It can be very easy for someone who doesn’t know that data to use it in a way it wasn’t intended for and be making the wrong interpretations of it.” [P4].

Despite data silos, most participants suggested that data sharing was not a particular issue (unless regulatory restrictions or GDPR apply). Furthermore, there was an apparent enthusiasm for innovation and change when discussing data sharing and transferring of data. A summary discussion around data ontologies and frameworks for shareable and reusable knowledge provoked a positive reaction from all participants, who were open to the idea of new ways to improve collaborative working and data sharing. More focus should be placed on establishing a data management protocol, where data can be stored and shared easily. Big data techniques can help with these challenges, especially when holding and processing several different data formats, but the approach should be intuitive and easily accessible (Ponce Romero et al., 2017).

7.5 Limitations of Pipe Failure Data

7.5.1 Coarse Recording the Time of Failure

Irrespective of the challenges posed from historical data and absent variables, readily available information reveals problems that limit pipe failure models, and of these, the lag time between the failure occurring and the failure being observed is a significant one. Sometimes pipe failures can persist underground unnoticed for some time before being discovered. Kleiner and Rajani (2012) highlighted the lag between actual failures and the time they are discovered, noting the effect of this on short-term predictions due to the momentary time-related conditions being obscured. Participants expressed difficulty with finding pipe failures which occur as small leaks in the network, acknowledging that some can go unnoticed for extended periods:

“Some failures have gone unnoticed for years, but sometimes we don’t know how long.” [P7].

“If it were days or weeks people would be very happy, I’ve known them to go [undetected] for years.” [P6].

Pipe failures can be difficult to detect, especially in certain conditions, for example where sandy soils are present, the water leaking from the pipe is less likely to reach the surface. Likewise, WDNs with a smaller percentage of metallic pipes are more difficult to listen for leaks in, because plastic pipes are not as receptive to the acoustic loggers used to detect failures.

The uncertainty of failure timing is inherent within most pipe failure datasets and requires a concerted effort by the water company to record the time of failure accurately. The authors have worked with a dataset that abates this problem to a certain extent by defining pipe failures as either 'reactive' or 'proactive'. Reactive pipe failures signify water emergence at the ground surface, posing a potential impact on continuous water supply, eliciting a quick response by the repair team (typically responding no later than 72 hrs). By this definition, failure time recorded is likely to be more accurate. However, it is also worth noting that water emerging at the surface may also be delayed by some time or even go completely unnoticed in rural areas. Conversely, proactive failures represent low-level leaks that may not immediately affect the network's operation, and as water remains below the surface, it often goes undetected. Proactive failures are actively sought using detection technicians and can be discovered long after the original failure incident.

The failure discovery lag time is a discernible problem, but summarising data by extended prediction intervals (e.g., annual, or greater) is one way to reduce the impact. However, if short-term predictions are required, then recording the time of failure accurately is necessary, and yet acknowledged as a considerable challenge. Acoustic sensors record real-time data and are used for detecting leakage more accurately. Using these data can help to accurately identify the time of failure (Water & Wastewater Treatment, 2017).

7.5.2 A lack of High-Resolution Data

Typically, data correlated with pipe failures originate from various sources at various scales (Kabir et al., 2016). Environmental data, such as weather or soil,

are just a few examples of secondary data. These data are estimated by area and, for weather, by time intervals under certain assumptions (Robles-Velasco et al., 2020), and therefore, do not capture the nature of the exacting conditions causing a failure event. For example, the Met Office Meteorological Office Rainfall and Evaporation Calculation Systems (MORECS) is apportioned over a 40 x 40 grid and summarised weekly. This, coupled with the lag times of changing weather conditions to the repercussions in soil movement and poorly located pipe failures (see Section 7.6.2), can obfuscate the exact conditions of failure (Kleiner and Rajani, 2012).

Another example of high-resolution data are localised soil conditions that are particularly important and often altered during excavation and construction (see Figure 7-2), this localised change in soil is not typically recorded. PVC pipes depend on lateral support from undisturbed soil (soil unmoved during the excavation, laying and backfilling of the pipe), on either side of the pipe to carry the backfill material load. Otherwise, the pipe can deform and flatten into an oval shape, reducing its resilience (Saadeldin et al., 2015). Since limited information is available regarding localised soil conditions the number of potentially flattened pipes cannot be quantified. Kleiner and Rajani (2012) noted a general lack of such geographically related data, just as with the experience of some participants:

“If you’ve got a national soil map on a large scale, the named ground you’re going to get in an urban environment, particularly one where it’s congested with other utilities, and you’ve got imported backfill mixed in with your topsoil, it can be hard to estimate the soil conditions around the pipe. This affects corrosion rates and shrink well etc.” [P4].

More localised data is required to improve failure predictions and understand the impact of localised weather and soil conditions. With emerging smart sensor technologies, detailed data collected at a higher spatial (individual pipes) and temporal (real-time) resolution for changes in soil conditions or weather conditions is possible and should intuitively improve failure predictions. Other

high-resolution data may also include pressure data, water temperature or water consumption for example. Yet, conditions for pipe installations should be documented in more detail to help understand localised soil conditions. These data, if available, may sometimes help explain variations in pipe failures.



Figure 7-2 Pipe installations showing the removal of local soil conditions, and installation of pipe bedding and backfill (Photos taken by author).

7.5.3 The Absence of Decommissioned Pipes

Pipe failure records are ordinarily limited compared to the life of the pipe, unless those assets are relatively modern (Economou et al., 2012). Consequently, pipes installed before failure records began may have failed numerous times but were never recorded, resulting in left-truncated data. Such data is present in most historical datasets as acknowledged by all participants, with one noting *“You’ve got the failure records which will be incomplete and do not go back to the start of the asset in most cases”* [P1] (Figure 7-3). It is important to emphasise that left-truncated data is not associated with left-censored data, a term sometimes used in pipe failure literature. Left-censoring only occurs if the number of breaks before pipe failure records is known, but when they happened is unknown, this is fundamentally different (Mailhot et al., 2000). Data collected during the failure

record period often remove discontinued pipe information to reflect latter network conditions (Figure 7-3), since as one participant noted, “*Decommissioned data is unnecessary for day-to-day operations.*” [P2].

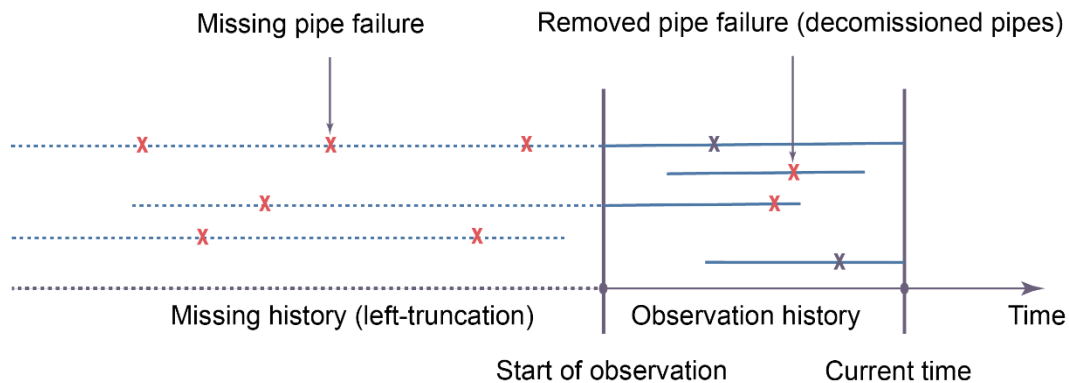


Figure 7-3 Depiction of pipe failure data with missing historical data (left truncation), removed pipe failures (decommissioned pipes), and reported failures.

Subsequently, helpful information about pipe failures is removed from the historical training dataset, information that could potentially improve the power of the model during training by increasing the failure sample size (Scheidegger et al., 2015). Nishiyama and Fillion (2013) suggest models utilising small data sets and few pipe failures regularly lead to low coefficients of determination and poor accuracy during a review of several studies. Many pipe failure models, such as deterministic models, do not account for these data characteristics (Scheidegger et al., 2015). Better accuracy has been reported when datasets contain more pipe failures and span some ≥ 30 years (Liu et al., 2012; Snider and McBean, 2019; Winkler et al., 2018). Decommissioned data should be retained and included in training datasets, yet data collection periods can only increase with time, which provides some optimism for improving future model accuracy.

7.6 Complexities of Pre-Processing

7.6.1 Handling Poor Quality Data

Pipe failure records are collected by repair teams who are under pressure to repair pipes quickly. Interruption due to data collection can lead to further costs and delays to customer service, and as one participant suggested, *“It’s not always important to point out a poorly recorded pipe age when the primary focus is to stop water from pouring out of the ground.”* [P3]. As a result, data collection can be inconsistent, incomplete, or inaccurate, as one participant acknowledged, *“You’ve got the failure records which will be incomplete and do not go back to the start of the asset in most cases. The records that we do have will be incomplete and have inaccuracies.”* [P4]. Poor data also results from equipment failure such as old, uncalibrated sensors creating signal noise (see Wu and Liu, (2017, p. 977) for further descriptions), from historical data and through human data processing, resulting in duplicate or erroneous data. Water companies should encourage complete data capture in the first instance and, where possible, should move to repair equipment failure quickly so as to ensure continuous data collection. This could be achieved by preparing data collection protocols and training staff on the importance of accurate and complete data collection, with one participant noting that *“Data collection is the biggest issue in that it is not done with the full understanding of how valuable the data is.”* [P6].

Despite attempts to improve data quality, some data quality issues will be present; therefore, important pre-processing steps are often required. Pre-processing approaches to address missing data are important since this can pose a negative impact on the use of data-driven approaches, leading to insufficient historical data to implement the pipe failure method (Wu and Liu, 2017). In instances where missing data is present a list-wise deletion (removing missing values) is often the easiest path to creating a workable dataset (Snider and McBean, 2019; Winkler et al., 2018). Yet discarding data results in a loss of information, introduction of bias, unreliable parameter estimates, underpowered models and poor convergence (Tang et al., 2019). Therefore, it is essential to

retain as much data as possible. Alternative approaches that maintain data include making inferences from event narratives or similarities with other events, especially for qualitative information such as material type or failure type. Here, errors in the interpretation can represent another source of uncertainty, for example,

“Typically, you would record either ‘cut out’ or ‘repair’, and you could infer a failure type from that, but it would certainly be unreliable...you won’t necessarily know why it’s failed. In fact, you probably don’t know why it’s failed.” [P4].

Alternatively, missing data can be handled through a range of approaches, imputing values based on probability functions or distributions from available datasets (Sattar et al., 2016), substituting mean values or standard deviation or through K-nearest neighbour (KNN) techniques, that calculate a value based on similarity to values within close proximity (Levinas et al., 2021). Machine learning methods also provide further solution, where some methods such as gradient boosting decision trees can accommodate missing data by imputing averages at the leaf node (Hastie et al., 2009). Yet approximations can obfuscate the exact localised conditions associated with pipe failure, introducing further uncertainty, and if a large amount of data is missing (typically > 25 %), then imputing can lead to incorrect results. Whilst these techniques aim to improve the data, pre-processing techniques are also available to reduce data complexity.

Data engineering processes such as feature extraction techniques use original data to obtain a new set of less redundant variables, optimising the model in the process (García et al., 2016). Further developments include feature engineering, a means of deriving new variables from domain knowledge. Recently, deep feature synthesis, has focused on automating new variables from existing relational data, providing new variables that perform well regardless of the model used (Rahbaralam et al., 2020). Data fusion techniques aggregate data from multiple sources to achieve improved, higher quality and more relevant information (Castanedo, 2013). Other popular feature engineering approaches exist such as one-hot-encoding, a procedure that converts original categorical

data into new binary values that are easier to interpret and are even necessary for some machine learning models (Aslani et al., 2021). Finally, many models have a high sensitivity to variables on different scales, especially so with some machine learning approaches. Standardisation can then be used to scale variables and reduce the effect of outliers that are often observed (Robles-Velasco et al., 2021).

7.6.2 Processing Incorrectly Placed Pipe Failures

Pipe failures are sometimes incorrectly located and appear some distance from any known pipe. Moreover, it is not uncommon for correctly located pipe failures to contain contrasting information with that already present in the existing pipe database (pipe diameter or material type are commonly incorrect).

“You rely on the contractor providing the right information back, and normally they can get the information right. But in all the mud and swamp of digging out the ground, they’ll not always get the pipe material or diameter size right because they are plastered, so it’ll be done afterwards.” [P6].

“Often guys in the field are under pressure to get the repair finished, especially if it is disrupting customers, so collecting data is not always a priority.” [P1].

“They don’t have enough people, and when they call incidents that need to get out and sort them, [data collection] it’s definitely not a priority.” [P8].

Since repair technicians train to understand pipe characteristics, contrasting information is presumed to be a result of incorrectly located pipe failures, where data has been updated after the repair event, or where data has been poorly recorded:

“Historically we won’t have included which particular pipe asset has failed [or] what the address was nearby, so if you got multiple assets in a street, you won’t know which one failed.” [P4].

“Location information from the original call is not always updated, especially if the inspection team are familiar with the area.” [P6].

Poorly located pipe failures are ordinarily handled during the data processing by either listwise deletion or relocation. As noted, listwise deletion can cause excessive data removal and subsequent problems; therefore, relocation is preferred. The authors have previously worked with data where some of the pipe failures required relocating to the nearest point on a pipe with similar characteristics, following a matching protocol firstly within 3 m, then 30 m, and then up to 1 km (see Figure 7-4). For pipe failures where information contrasts the pipe database, this method provides a valuable solution yet presumes that both sets of data are correct. Conceivably, this may not be the case, and if not, the error persists and can thus mostly pass undetected.

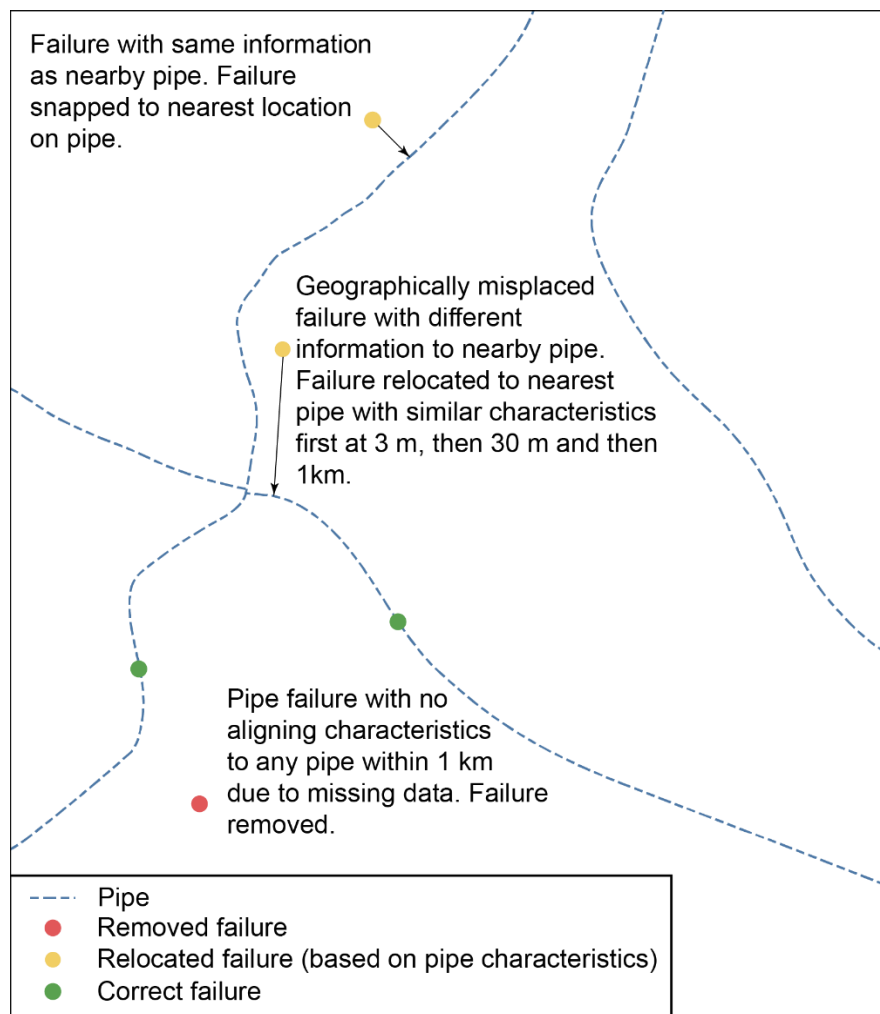


Figure 7-4 Illustrative example of a water distribution network with methods for the relocation of pipe failure data. Green – correctly geographically located pipe failures snapped to nearest pipe. Yellow – geographically misplaced pipe failure moved to the nearest pipe within 1 km. Red – failure shares no characteristics of any pipe within 1 km.

Understandably, pipe failures are difficult to accurately predict when the location of several pipe failures remains in question. Water companies must ensure data collection in the field is accurate by checking details against those records within the existing pipe database. Thus, by highlighting any contrasting information during excavation, there is a chance to resolve any data issues before the repair has finished, an approach likely to reduce the number of failures that need relocating. Likewise, pipe failures should be correctly located with the pipe

network where the failure occurred. One participant noted that creating easy-to-use software such as drop-down boxes with specific responses is not always beneficial: *“We used dropdown boxes to make it easier, but the guys tend to always go for the first option regardless. The same thing was observed by many water companies.”* [P6]. Incentivising and encouraging data capture, highlighting its importance, or auditing data capture may be more appropriate. However, organisational silos where teams work specifically towards their own business needs can be a challenge. It is important to encourage communication and collaborative working between teams so that the wider company goals are achieved.

7.6.3 Using Pipe Length and Previous Failures

Pipe length is a proxy for exposure to failure, where longer lengths denote a higher risk, and the number of failures is directly proportional to the pipe length (Kleiner and Rajani, 2012; Robles-Velasco et al., 2020). In studies of failure rate, the pipe length, used as a log offset, normalises the failure rate and implies a uniformly distributed number of failures along the length of the pipe (Kleiner and Rajani, 2012). Other studies have used pipe length as a covariate to explain failures, believing this is a more appropriate approach (Shamir and Howard, 1979; Winkler et al., 2018) or even as a unit for determining other variable levels (Robles-Velasco et al., 2020). Kleiner and Rajani (2012) discussed the relationship between pipe failures and pipe length, reporting a significant element of randomness. However, a relationship was present when summarising groups of pipes with similar characteristics, accumulating enough pipe failures and length for this randomness to reduce. Pipe length can also have a statistical meaning for other related factors, such as traffic loading, operational stress or bedding conditions, yet this is impossible to quantify (Berardi et al., 2008). Pipe length is important, yet it can be problematic since, in many networks, considerably disparate lengths of pipes have been identified, making them difficult to model. In this regard, some authors have re-cut the lengths of pipe by street to get more accurate predictions (Winkler et al., 2018). Alternatively, Berardi et al. (2008)

suggest that all the variables related to pipe length should be represented by length related mean, whilst others propose logarithmically transforming the data (Robles-Velasco et al., 2021). In models where pipes are grouped, they are typically done so by available soil characteristics and fail to represent similar conditions for traffic loading and operational stress, amongst others.

Failure models use pipe failure history explicitly or implicitly. Implicitly, pipe failure history is used as a surrogate for deterioration and the performance of the pipe, and is represented as either the total number of previous failures or transformed and used as a covariate or stratification criterion. Kleiner and Rajani (2012) used recency (a measure of how recent the last failure occurred) as an expression of pipe failure history, which considers whether historical breaks have occurred more at the beginning or end of the observation period. However, as previously discussed, most pipe failure datasets are prone to left truncation, where failures, before the records started, are absent, resulting in an unreliable estimate of deterioration for a given pipe. Pipe age is also used to measure deterioration, suggesting that older pipes fail more often since they have been subject to deterioration over a more extended period. Again, many participants identified several occasions where older pipes outperformed new pipe installations, with one participant adding “*Age is different to performance and shouldn’t be considered because old pipes often outperform new pipes.*” [P2]. The type of model used should dictate how and if failure history can be used as a covariate, but in the absence of knowledge on the pipe condition, the number of previous failures could be a valuable surrogate for deterioration and resilience to failure.

7.7 Understanding Model Scalability

7.7.1 Scaling Geographic Extent

Typically, water companies would prefer failure models to report on individual assets, yet for individual pipes, failures are infrequent. Consequently, the combination of infrequent failures and short failure records mean datasets for individual pipe models are imbalanced (more non-failures than failures present

with a ratio $> 9:1$ ratio). Three common ways to address data imbalances reported from the literature include the following:

- Scaling the model to either use longer prediction intervals (greater than annual failures) or grouping pipes by similar characteristics to accumulate more failures by the length of pipe.
- For classification models, adopting the approach of over and under-sampling, class weighting (Winkler et al., 2018) or Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002). For regression models, the use of zero-inflated models has been reported as effective (Economou et al., 2012). However, there are limitations to sampling methods and zero-inflation models, and the results are not always favourable, nor do they necessarily improve the model accuracy (Kleiner and Rajani, 2012). These techniques are not always necessary for probabilistic modelling.
- Increasing failure records and ensuring decommissioned data remains in the data set, *'The best way to tackle this problem is always to collect more data.'* (Chicco, 2017, p. 7).

Predicting failures at the asset level is seen as desirable since many management decisions are made for individual pipes, yet these models have typically resulted in poor accuracy (Liu et al., 2012). This was also observed by participants who have experience of industry working failure models, with one saying, *"It would be lovely to have a model running at the asset level, but I don't think there's any company that has sufficient granularity and accuracy in their data to deliver that."* [P6]. A desirable and common precursor to pipe failure analysis is partitioning pipes into homogenous groups based on similar characteristics, or in some recent studies, by using k-means clustering (Kleiner and Rajani, 2012). Grouping pipes by similar characteristics whilst summing the number of pipe failures and pipe length increases the number of failures and reduces data imbalance for the model to train on, resulting in greater accuracy (Kleiner and Rajani, 2012). Kleiner and Rajani (2012) concluded that inherent

uncertainty and a lack of data makes individual pipe failure models unreliable, which some participants also suggested, *“For smaller diameter pipes, we found that creating cohorts is necessary to make any appreciable progress.”* [P4]. However, grouping can amass pipe lengths several km long, which does not always aid localised management decisions, but can highlight potential areas for further investigation. Grouping pipes also assumes that all pipes within the group share similar failure rates, which is not often the case. In particular, one pipe may have a far greater number of failures than others within the group, consequently increasing the failure rate for all pipes unnecessarily.

7.7.2 Scaling Dynamic Time-Dependent Variables

Early failure models traditionally represented the steady deterioration of pipes. As models evolved, researchers started to include dynamic time-dependent variables, yet mathematically, the statistical analysis is substantially challenging. Dynamic time-dependent variables can be variable or cyclical, and they can often mask the effects of pipe deterioration. The effectiveness of using these variables relies upon the time interval used in forecasting them accurately into the future. To accommodate the fluctuations in seasonal variation, predictions intervals should be less than annual, typically interannual or monthly, yet decreasing the time interval increases the number of observations in a dataset to a point where the data imbalance naturally becomes difficult to compute with any accuracy. Some participants were more sceptical regarding short-term failures due to the experience gained within their own companies, with one participant stating *“I certainly wouldn’t be placing much faith on these models on a weekly or monthly basis.”* [P4]. The use of weather covariates is typically within the training model only and cannot be used in forecasting unless weather forecast data is included. Forecasting weather data accurately is a complex task, especially with global temperature change, and would add another dimension of uncertainty to the final predictions. However, weather covariates are used since their inclusion is likely to reveal the *‘True background ageing rates’* through showing increased failure frequency over time (Kleiner and Rajani, 2012, p. 671).

7.8 Future Opportunities and Challenges

Informed proactive management decisions seek to determine the right place and time for active intervention. Big data collected through smart sensors provides an opportunity to improve the accuracy of predictions by providing continuous high-resolution real-time data capture (such as water pressure data) and a wider variety of variables associated with the mechanisms of failure. This was discussed by participants, one suggesting *“We are on a journey, and at some point, in the near future data will be available to improve models.”* [P5]. With technology driving big data collection, actionable information will start to emerge, and it is feasible that water companies will find alternative means to reduce pipe failures. For example, ontologies using DAS deployed across the network records real-time data that can detect unique failure characteristics using a statistical model to detect specific acoustic signatures, alerting users to a leakage (Levinas et al., 2021). This approach can detect a leak within 10 metres of its location after only 30 litres of water has been lost (Water Industry Journal, 2018), providing a real-time accuracy that statistical models alone have not yet achieved.

Recent focus has been placed on analytics, where automated decisions are built based on real-time data, reducing the amount of water lost by detecting pipe failures and reacting with an appropriate course of action (e.g., reducing pipe pressure) (Lin et al., 2012). One participant noted, *“I’ve seen a lot of great work from our data science team, and I think we are on the cusp of understanding and realising a world of real-time modelling.”* [P7]. Digital twin approaches extend real-time modelling by providing a virtual representation of an asset, spanning its life cycle, using real-time data, machine learning and simulation to help decision-making. Digital twins are being explored for leakage, energy efficiency, water quality and maintenance planning to name a few, but will be an essential support system for future management (Conejos Fuertes et al., 2020). Using smart sensors, operation teams can respond quickly to failures, reducing service interruptions (Water & Wastewater Treatment, 2017). It is unlikely that failure models alone will compete with smart networks for operations teams in the near

future, especially when short-term predictions for short sections of a pipe typically reveal poor accuracy (Kleiner and Rajani, 2012). However, failure models can still be used to inform long-term asset management decisions, identifying areas for rehabilitation or replacement. By shifting focus towards proactive management, water utilities can reduce cost and inefficiencies and build future resilience into the network. Yet failure models need to provide a level of accuracy that gives confidence to the predictions:

“It’s always challenging, and whatever you predict is never going to be exactly right...you will always have a problem convincing people that your model is right because it probably isn’t.” [P4].

In this respect, it is perhaps more helpful to predict the probability of failure or contribute risk and vulnerability maps that can rank pipes against each other in relative terms instead of predicting the exact number of failures. Such approaches can even provide helpful insights at an engineering level to take constructive action towards extending life expectancy and building future resilience into the network. Targeting areas of the network this way can also help to facilitate teams actively searching for leaks that are difficult to find, locating optimal areas in which to deploy hydrophones and ground microphones.

There is a widespread belief that machine learning provides an opportunity to augment more traditional statistical models and improve predictions since they can extract complex relationships (Snider and McBean, 2019). However, for limited structured data, often seen in pipe failure datasets, the improvements are often minor after data pre-processing is conducted. Given the additional development and computational time, the most straightforward approaches may prove the most effective. Simple models offer an easier way to explore the predicted outputs and provide interpretability, albeit only in the hands of professionals with domain specific knowledge (Rudin, 2019), with one participant noting, *“As far as possible we try not to have a completely black-box approach that relies heavily on a kind of machine learning where you can’t easily work out how you got from your start point to your endpoint.” [P4].* Yet there is wide interest

in machine learning, and many models have shown improved accuracy. However, these more advanced methods cannot significantly improve pipe failure predictions based on insufficient data. One participant noted *“The challenge with this is that the insight is only ever going to be as good as the data.”* [P3]. Therefore, it is likely that a combination of improved data and more advanced models can help support the rationale for an improved decision-making process. Large complex datasets with many variables and complex relationships are where machine learning models perform best, but this also depends on the ability of the researcher to create accurate, interpretable models (Rudin, 2019).

7.9 Operational Practicality and Methodology of the Approach developed within this research

To gauge the operational practicality of the pipe failure methodologies, it is necessary to elaborate the approaches in discussion with experienced senior industry professionals, eliciting opinions as to its potential use. The discussion was conducted after a presentation of the modelling approach, with deliberation conducted across four main areas: 1) operational applicability of the model, 2) practicality of the spatiotemporal scale, 3) variable selection, and 3) model accuracy. The results are presented hereafter.

7.9.1 Operational Applicability of the Model

The model was developed to help asset management decisions. A presentation regarding the modelling approach adopted was provided and participants then asked where such a model might be utilised within the company to understand its initial use and possible use within the wider sector (e.g., operational). Several responses were elicited:

“Definitely for asset management, but the ops guys won’t use this” [P01]

“This looks similar to the model we do [use] for strategic purposes, for planning what investment in our pipes were going to be making over the next AMP period, and to a certain degree for longer term planning. The approach we have works

adequately well, but there is scope to look at different methods for larger pipes.”
[P3]

“This model doesn’t help operational teams, or prevent bursts, but it does help with asset costing.” [P5]

“My immediate thoughts are whilst it initially sits in asset management function is something we all need to be aware of and from a risk management perspective. I think outside of asset management there are a lot of people that are monitoring the network so whether it be in leakage or optimisation, colleagues that are looking for opportunities will be looking for models like this.” [P8]

From the responses received, it was determined that modelling approaches such as these could offer a useful approach for other parts of the business, such as within the leakage team, for identifying areas of the network within which to prioritise a search for leaks, or for future costing and budget planning purposes. From the discussions, the operational team indicated a reluctance to adopt such modelling approaches unless the scale of the model could be improved, with one participant noting, *“For operational teams, your model would have to say your pipe is going to fail at this time on this day for them to have any interest, but it is good for mid-level planning.”* [P6]. Operational teams typically require short-term predictions to provide early warning or real-time losses (Jenkins et al., 2018). However, DAS deployed across the network offers a more competitive solution, whereby a leak can be detected within 10 metres of its origin (Levinas et al., 2021).

7.9.2 Practicality of the Spatiotemporal Scale

The participants were asked their views regarding the usefulness of the spatiotemporal scale adopted. It was highlighted that the model was based upon groups of pipes aggregated by a 1 km grid, and the predictions made were provided on an annual basis. The participants had mixed views regarding the model scale depending on the experience within their company. Some participants viewed the pipe groupings as essential, with one noting, *“It’s a*

challenging statistical problem because of the sparsity of failures. Creating groups of pipes can improve accuracy.” [P6]. Another participant indicated that asset level predictions were more useful, but also noted their understanding of the difficulty when providing accurate predictions at this level, suggesting, *“It would be more useful to have the model running at the asset level, but I suppose the results wouldn’t be much use.”* [P8].

Two participants further discussed the potential for predicting failures for DMAs. One participant suggested that pipe replacement be performed for whole DMAs that are underperforming since the cost of mobilising and excavating can be high; therefore, working systematically in a single area is financially beneficial, *“The overall approach is ok, but did you consider looking at the performance of DMA’s? We usually look at DMA performance”* [P2]. In this research, grouping by DMA was considered, yet the pipe groups were too small at this level for meaningful results to be attained on an annual scale; therefore, the 1 km grid approach was used. However, predicting pipe failures at the DMA level would clearly be helpful, and for longer prediction intervals such as AMP periods (five years), this may be useful. Predicting failures in DMAs is something that merits further investigation.

7.9.3 Variable Selection

The participants were asked whether the variables chosen were seen as factors that they might consider influential to pipe failures and whether anything else should have been included in addition. Most of the variables used were seen as necessary, with only age of pipe being questioned, with one participant suggesting, *“Age and performance are different, so age shouldn’t be considered, because old pipes often outperform new pipes.”* [P4]. Whilst this is true to an extent, age was established as a statistically significant variable in this study, and has been further noted as necessary in most studies (Barton et al., 2022).

The inclusion of pipe pressure was recommended by most participants, with one suggesting *“Hydraulic parameters should be included. A good cover of flow rates and pressure is usually available.”* [P1]. Pipe pressure is known to be important

(Robles-Velasco et al., 2020; Yamijala et al., 2009); however, as previously discussed, pipe pressure was unavailable for this study.

Other notable comments include the use of smart meter information and the potential influence of commercial users, suggesting, *“You should have a look at non-residential users and commercial metres; I don’t see this in your work.”* [P7]. The influence of large users was explored in Chapter 2 but without any indication of influence on failure rates. However, further studies on the influence of commercial users may well prove to be helpful.

One participant made an important observation on the geographical scale of some of the data, suggesting *“The variable scales probably don’t reflect local conditions.”* [P2]. Data scale has been previously identified as a limitation, and further discussion on data granularity and real-time data can be found in Section 7.5.2.

7.9.4 Model accuracy

The accuracy of pipe failure models is a point of contention since several studies have reported poor accuracy, typically where models are built on a smaller spatial (asset level) or temporal (less than annual) scale. However, smaller spatial scales are the most useful to water companies, thus a balance must be struck between the usefulness of the model and the level of accuracy. The discussion of model accuracy is an important one and understanding the perceptions of accuracy is important. When asked, some of the participants suggested that management would impose a particular level of accuracy, with one noting, *“I wouldn’t go on accuracy, but management are obsessed with 75 % accuracy, this level of accuracy seems arbitrary, but anything less and they won’t accept it.”* [P1], and another saying, *“I should think you would need to get above 60 % accuracy for anyone to even look at it.”* [P5]. Most of the participants understood the value of the model and took a more pragmatic approach to assess the level of accuracy, with two participants noting:

“There are lots of ways of looking at how good a model is, and different ways will make different models look better or worse. I think it can be dangerous to quote something that seems quite high, it’s more important to think about how actionable the outputs of the model are.” [P3]

“For me, it’s all around the so what. What are we going to do with it, how are we going to use it? The point of accuracy is an important one, but really, the model needs to be useful.” [P8]

This is important since there are several accuracy measures, and the literature has reported that accuracy is not always a useful gauge of the value of the predictions. This is especially evident where imbalanced data is present, such that predicting all non-failures correctly would lead to very high accuracy, even if no failures were correctly predicted (Barton et al., 2021). From the discussions, it was noticeable that participants understood the limitations of models and did not expect to see results that could replicate the real world, with one person suggesting, *“if the model accuracy was very high, I would be very concerned that something was wrong”* [P2]. One of the participants suggested splitting pipes into large and small diameters to improve the predictive accuracy of the model *“We have found splitting the larger diameter pipes and smaller diameter pipes more useful.”* [P3], suggesting larger diameter pipes were more predictive when using a hybrid of statistical and physical modelling.

A recent development in pipe failure model literature is the use of rank aggregation to help evaluate models on their accuracy to predict higher or lower failure rates rather than accurately predicting the failure rate itself. In this respect, models should be judged on their capability to rank pipes, quantified by counting the total failures avoided (Aslani et al., 2021; Chen and Guikema, 2020).

7.9.5 Comments

The discussions held with industry professionals suggest that the methodology was viewed as potentially useful for asset management planning, its primary purpose. In partnership with expert knowledge, the participants could clearly see

how the pipe ranking approach could support decision makers. All participants questioned were able to discuss the limitations, and from this discuss potential areas for further improvement. From this it can be concluded that future research would benefit from exploring failures at the DMA level or segmentation of the pipes in alternative ways to improve accuracy, seeking more accurate local data and including more relevant variables.

Despite this, the participants further suggested having ‘some’ information on future failures is better than current approaches which do not include any future predictions.

7.10 Key Findings

To highlight important aspects of the discussion, the key findings are summarised and presented in Table 7-2, separated by each of the five themes.

Table 7-2 Summary of key findings including challenges and priorities for future efforts to improve pipe failure models.

Challenge	Findings
Challenges of Managing Data	<ul style="list-style-type: none">▪ Collect continuous data across regulatory periods.▪ Collect more data correlated with pipe failures in useable formats.▪ Remove data silos and promote the use of ontological data sharing.▪ Establish data management protocols, and harmonise the disparate databases, providing data polls that are initiative.
Limitations of Pipe Failure Data	<ul style="list-style-type: none">▪ Summarise prediction intervals (e.g., annual, or greater) to reduce the effects of coarsely recorded failure times.▪ Use real-time data collection if available to estimate the time of failure (e.g., pressure or flow data).▪ Determine reactive and proactive failures.▪ Use high-resolution data using smart sensors.▪ Collect information on localised conditions during excavation operations to repair and lay new pipes.▪ Use decommissioned pipe data for training data sets.

Challenge	Findings
Complexities of Pre-processing	<ul style="list-style-type: none"> ▪ Communicate the importance of data to encourage accurate and complete repair data capture – especially location. ▪ Establish standardised formal data collection protocols. ▪ Use advanced pre-processing methods such as imputing, KNN, feature engineering or data fusion. ▪ Devise a suitable means to place pipe failures accurately, linking site records to pipe records. ▪ Consider dissimilar asset lengths by grouping, re-cutting, length related mean or logarithmic transformation. ▪ The type of model used should dictate how and if failure history can be used as a covariate.
Understanding model scalability	<ul style="list-style-type: none"> ▪ The desired output should determine the model scalability, although short-term predictions for individual pipes perform poorly. ▪ At low spatial and temporal scales, data imbalanced must be carefully considered, perhaps using sampling techniques or zero-inflated models.
Future opportunities and challenges	<ul style="list-style-type: none"> ▪ Collection of improved quality data must be prioritised. ▪ Advanced data-driven techniques such as machine learning, used in combination with continuous data offers more accurate predictions. ▪ The use of digital twins, smart sensors and real time data offers more accuracy than short-term pipe failure models for operations teams. ▪ Failure models can be used for long-term strategy management.

7.11 Study Limitations

There are several difficulties with conducting semi-structured interviews. Ensuring respondents understand the question can be problematic, primarily when broad and open-ended (Oltmann, 2016). Addressing this, during the interviews, the initial questions were well understood as only practitioner participants with a depth of professional knowledge were selected, their experience, leading to further interesting discussions. Also, responses can be

biased since they are the opinions of the individuals, and whilst the professionals involved provided thorough responses, there may have been useful information that participants were not privy to or held back on. Bias can also be introduced by the interviewers verbal and non-verbal cues (Ritchie et al., 2003). Yet, questions were asked in a manner intended to reduce bias by refraining from becoming too involved in the discussion and offering comments (Adams, 2015). Face-to-face interviews are particularly prone to reactivity problems, where respondents provide socially acceptable responses. However, given the nature of the questions, which were not sensitive or controversial and the open ended approach, the likelihood of reactivity in responses is limited (Oltmann, 2016). This was further limited by turning the camera off during the interview, becoming a method of a phone call as suggested by Oltmann (2016). The information gathered is restricted in geographical scope, and in this respect, it is not generalisable.

7.12 Conclusion

This study has sought to discuss the key challenges and uncertainty in pipe failure models holistically, drawing from the literature and complementing the findings with further understanding from current practices in water companies. The findings reveal that uncertainty stems from poor data quality and quantity, resulting from challenges presented when collecting and managing the data. There needs to be a focus on data collection, in particular communication as to the importance of the data, continuous data collection, sharing relevant data and collecting continuous data. Improving data in the first instance will ensure reliable pipe failure models. Further challenges arise from the infrequent number of failures in pipe failure networks that present mathematical probable to model. The combination of this and poor data quality and quantity makes short-term-modelling (weekly, or monthly) for individual assets difficult, and is unlikely to provide an accuracy that guides management decisions.

Modelling pipe failure is an essential component in the proactive management of assets, enabling the identification of areas in the WDN most at risk of failure, guiding replacement strategies, rehabilitation and maintenance schemes or helping target areas of the network for further leakage investigation. With the advent of smart networks and improving analytical techniques, water companies can be more proactive in managing pipe failures, using real-time modelling and insight. The collection of vast data and potentially using modern analytics and machine learning methods offers information to improve pipe failure models, helping water companies make better decisions, especially considering future water demands and a changing climate.

7.13 Author Contributions

Neal Barton: Conceptualisation, Methodology, Investigation, Data Curation, Writing – Original Draft, Writing – Review and Editing, Visualisation and Project administration. Stephen Hallett: Writing – Review and Editing, Supervision, Funding Acquisition. Simon Jude: Review and Editing, Supervision, Funding Acquisition.

7.14 References

- Adams, W.C., 2015. Conducting Semi-Structured Interviews, in: Newcomer, K.E., Hatry, H.P., Wholey, J.S. (Eds.), Handbook of Practical Program Evaluation. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 492–505. <https://doi.org/10.1002/9781119171386.ch19>
- Aslani, B., Mohebbi, S., Axthelm, H., 2021. Predictive analytics for water main breaks using spatiotemporal data. Urban Water J. 00, 1–16. <https://doi.org/10.1080/1573062X.2021.1893363>
- Barton, N.A., Farewell, T.S., Hallett, S.H., 2020. Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. npj Clean Water 3, 31. <https://doi.org/10.1038/s41545-020-0077-3>

- Barton, N.A., Farewell, T.S., Hallett, S.H., Acland, T.F., 2019. Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Res.* 164, 114926. <https://doi.org/10.1016/j.watres.2019.114926>
- Bengtsson, M., 2016. How to plan and perform a qualitative study using content analysis. *NursingPlus Open* 2, 8–14. <https://doi.org/10.1016/j.npls.2016.01.001>
- Berardi, L., Giustolisi, O., Kapelan, Z., Savic, D.A., 2008. Development of pipe deterioration models for water distribution systems using EPR. *J. Hydroinformatics* 10, 113–126. <https://doi.org/10.2166/hydro.2008.012>
- Castanedo, F., 2013. A review of data fusion techniques. *Sci. World J.* 2013. <https://doi.org/10.1155/2013/704504>
- Chawla, N. V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* 16, 321–357. <https://doi.org/10.1613/jair.953>
- Chen, T.Y.-J., Beekman, J.A., David Guikema, S., Shashaani, S., 2019. Statistical Modeling in Absence of System Specific Data: Exploratory Empirical Analysis for Prediction of Water Main Breaks. *J. Infrastruct. Syst.* 25, 04019009. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000482](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000482)
- Chicco, D., 2017. Ten quick tips for machine learning in computational biology. *BioData Min.* 10, 35. <https://doi.org/10.1186/s13040-017-0155-3>
- Conejos Fuertes, P., Martínez Alzamora, F., Hervás Carot, M., Alonso Campos, J.C., 2020. Building and exploiting a Digital Twin for the management of drinking water distribution networks. *Urban Water J.* 17, 704–713. <https://doi.org/10.1080/1573062X.2020.1771382>

- Economou, T., Kapelan, Z., Bailey, T.C., 2012. On the prediction of underground water pipe failures: zero inflation and pipe-specific effects. *J. Hydroinformatics* 14, 872–883. <https://doi.org/10.2166/hydro.2012.144>
- Fusch, P., Ness, L., 2015. Are We There Yet? Data Saturation in Qualitative Research. *Qual. Rep.* 20, 1408–1416. <https://doi.org/10.46743/2160-3715/2015.2281>
- Galvin, R., 2015. How many interviews are enough? Do qualitative interviews in building energy consumption research produce reliable knowledge? *J. Build. Eng.* 1, 2–12. <https://doi.org/10.1016/j.jobbe.2014.12.001>
- García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J.M., Herrera, F., 2016. Big data pre-processing: methods and prospects. *Big Data Anal.* 1, 1–22. <https://doi.org/10.1186/s41044-016-0014-0>
- Guest, G., Bunce, A., Johnson, L., 2006. How Many Interviews Are Enough? *Field methods* 18, 59–82. <https://doi.org/10.1177/1525822X05279903>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning*, Springer Series in Statistics, Springer Series in Statistics. Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-84858-7>
- Kabir, G., Tesfamariam, S., Loepky, J., Sadiq, R., 2016. Predicting water main failures: A Bayesian model updating approach. *Knowledge-Based Syst.* 110, 144–156. <https://doi.org/10.1016/j.knosys.2016.07.024>
- Kakoudakis, K., Farmani, R., Butler, D., 2018. Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. *J. Hydroinformatics* 20, 1191–1200. <https://doi.org/10.2166/hydro.2018.152>
- Kleiner, Y., Rajani, B., 2012. Comparison of four models to rank failure likelihood of individual pipes. *J. Hydroinformatics* 14, 659–681. <https://doi.org/10.2166/hydro.2011.029>

- Konstantinou, C., Stoianov, I., 2020. A comparative study of statistical and machine learning methods to infer causes of pipe breaks in water supply networks. *Urban Water J.* 17, 534–548. <https://doi.org/10.1080/1573062X.2020.1800758>
- Levinas, D., Perelman, G., Ostfeld, A., 2021. Water Leak Localization Using High-Resolution Pressure Sensors. *Water* 13, 591. <https://doi.org/10.3390/w13050591>
- Likhari, R., Webb, S., Bricker, S., Bonsor, H., Frith, N., Catchpole, S., 2017. Mapping Underground Assets in the UK 11–12.
- Lin, J., Sedigh, S., Hurson, A.R., 2012. Ontologies and Decision Support for Failure Mitigation in Intelligent Water Distribution Networks, in: 2012 45th Hawaii International Conference on System Sciences. IEEE, pp. 1187–1196. <https://doi.org/10.1109/HICSS.2012.458>
- Liu, Z., Kleiner, Y., Rajani, B., Wang, L., Condit, W., 2012. Condition assessment technologies for water transmission and distribution systems, United States Environmental Protection Agency.
- Mailhot, A., Pelletier, G., Noël, J.-F., Villeneuve, J.-P., 2000. Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application. *Water Resour. Res.* 36, 3053–3062. <https://doi.org/10.1029/2000WR900185>
- Nishiyama, M., Fillion, Y., 2013. Review of statistical water main break prediction models. *Can. J. Civ. Eng.* 40, 972–979. <https://doi.org/10.1139/cjce-2012-0424>
- Ofwat, 2020. Innovation in the water sector [WWW Document]. URL <https://www.Ofwat.gov.uk/regulated-companies/innovation-in-the-water-sector/>

- Ofwat, 2006. The Development of the Water Industry in England and Wales, Ofwat. Birmingham.
- Oltmann, S.M., 2016. Qualitative interviews: A methodological discussion of the interviewer and respondent contexts. *Forum Qual. Sozialforsch.* 17. <https://doi.org/10.17169/fqs-17.2.2551>
- Ponce Romero, J., Hallett, S., Jude, S., 2017. Leveraging Big Data Tools and Technologies: Addressing the Challenges of the Water Quality Sector. *Sustainability* 9, 2160. <https://doi.org/10.3390/su9122160>
- Rahbaralam, M., Modesto, D., Cardús, J., Abdollahi, A., Cucchiatti, F.M., 2020. Predictive Analytics for Water Asset Management: Machine Learning and Survival Analysis 1–19.
- Ramirez, R., Torres, D., López-Jimenez, P.A., Cobacho, R., 2020. A front-line and cost-effective model for the assessment of service life of network pipes. *Water (Switzerland)* 12, 1–23. <https://doi.org/10.3390/w12030667>
- Renaud, E., Massiac, J.C. De, Brémond, B., Laplaud, C., 2007. SIROCO, A Decision Support System for Rehabilitation Adapted for Small and Medium Size Water Distribution Companies, in: LESAM 2007 - 2nd Leading Edge Conference on Strategic Asset Management. International Water Association, Lisbon, Portugal, pp. 1–15.
- Ritchie, J., Lewis, J., Elam, G., 2003. Designing and Selecting Samples, in: Ritchie, J., Lewis, J. (Eds.), *Qualitative Research Practice*. SAGE Publications Inc, Trowbridge, Wiltshire, pp. 77–108.
- 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. *Reliab. Eng. Syst. Saf.* 196, 106754. <https://doi.org/10.1016/j.ress.2019.106754>

- Robles Velasco, A., Muñuzuri, J., Onieva, L., Rodríguez Palero, M., 2021. Trends and applications of machine learning in water supply networks management. *J. Ind. Eng. Manag.* 14, 45. <https://doi.org/10.3926/jiem.3280>
- Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* 1, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>
- Saadeldin, R., Hu, Y., Henni, A., 2015. Numerical analysis of buried pipes under field geo-environmental conditions. *Int. J. Geo-Engineering* 6. <https://doi.org/10.1186/s40703-015-0005-4>
- Sattar, A.M.A., Gharabaghi, B., McBean, E.A., 2016. Prediction of Timing of Watermain Failure Using Gene Expression Models. *Water Resour. Manag.* 30, 1635–1651. <https://doi.org/10.1007/s11269-016-1241-x>
- Scheidegger, A., Leitão, J.P., Scholten, L., 2015. Statistical failure models for water distribution pipes – A review from a unified perspective. *Water Res.* 83, 237–247. <https://doi.org/10.1016/j.watres.2015.06.027>
- Shamir, U., Howard, C.D.D., 1979. An Analytic Approach to Scheduling Pipe Replacement. *J. Am. Water Works Assoc.* 71, 248–258. <https://doi.org/10.1002/j.1551-8833.1979.tb04345.x>
- Snider, B., McBean, E.A., 2019. Improving Urban Water Security through Pipe-Break Prediction Models: Machine Learning or Survival Analysis. *J. Environ. Eng.* 146, 04019129. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001657](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001657)
- Sun, A.Y., Scanlon, B.R., 2019. How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. *Environ. Res. Lett.* 14, 073001. <https://doi.org/10.1088/1748-9326/ab1b7d>

- Tang, K., 2018. Statistical Modelling Approaches to Improve Risk Assessments for United Kingdom's Water Infrastructure. [unpublished] Cranf. Univ.
- Tang, K., Parsons, D.J., Jude, S., 2019. Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliab. Eng. Syst. Saf.* 186, 24–36. <https://doi.org/10.1016/j.ress.2019.02.001>
- UKWIR, 2020. The National Failure Database [WWW Document]. URL <https://ukwir.org/the-national-failure-database> (accessed 2.28.21)
- Water & Wastewater Treatment, 2017. Leakage: Acoustic loggers strike the right note for Affinity Water [WWW Document]. URL <https://wwtonline.co.uk/features/southern-water-s-leakage-reduction-drive> (accessed 10.6.20)
- Water Industry Journal, 2018. Reducing leaks: the rise of leak detection technology in water pipes [WWW Document]. URL <https://www.waterindustryjournal.co.uk/reducing-leaks-the-rise-of-leak-detection-technology-in-water-pipes> (accessed 10.7.20)
- Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., Tscheikner-Gratl, F., 2018. Pipe failure modelling for water distribution networks using boosted decision trees. *Struct. Infrastruct. Eng.* 14, 1402–1411. <https://doi.org/10.1080/15732479.2018.1443145>
- Wu, Y., Liu, S., 2017. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water J.* 14, 972–983. <https://doi.org/10.1080/1573062X.2017.1279191>

8 Summary of Findings and Implications of the Work

Objective 6	Chapter 8: Summary of findings and implications of the work	<ul style="list-style-type: none">▪ Summarise the research findings with respect to the aims and objectives and contributions.▪ Implications of the work, including examining the views of the proposed model as discussed with industry professionals.▪ Recommend further research
-------------	---	---

8.1 Introduction

This research has presented a body of work through an embedded case study design that has sought to understand, investigate, and predict the risk of pipe failures in a UK WDN. The following chapter synthesises the findings from the previous chapters, connecting the Research Objectives. The research findings presented in Chapters 2 through 7 present original work with new insights into the complexity of managing ageing infrastructure. The implications of the research have been discussed and recommendations for further work offered.

8.2 Summary of Findings and Contribution

The findings of this research respond to gaps in knowledge relating to advancing the understanding of predicting the risk of pipe failures for proactively managing water distribution networks. The findings are discussed by Objective and synthesise the key findings and wider end-user context.

Objective 1: Understand and characterise the modes and mechanisms of failure in WDNs.

Statistical based modelling relies on domain knowledge to identify appropriate variables likely to be correlated with pipe failures. Chapter 2 comprises a literature review focussing on factors affecting pipe failures, investigating the complexity of pipe failures by organising common factors that form the various modes and

mechanisms of pipe failure, categorised as Pipe-intrinsic, Environmental and Operational.

Pipe intrinsic. General observations from the literature suggest that pipe failures are more likely in older pipes (due to pipe age and manufacturing periods) and that these pipes are less resilient due to corrosion (only in iron, AC and SDI) and smaller pipes with thinner walls. Plastic pipes are less susceptible to corrosion but can fail due to manufacturing defects or poor material handling prior to installation, resulting in micro-cracks. Earlier manufacturing periods also seemingly cause higher failure rates in pipes. A further factor is whether the pipe has been protected by a coating or lining before installation or during the lifespan of the pipe. Rigid joint systems are found to be less resilient to ground movement than are push-fit joints.

Environmental. Iron, AC and SDI materials are prone to failures arising from the combined effects of extreme variations in weather conditions (e.g., temperature, frost and SMD) that influence ground movement, and from the aggressive clay and peat soil conditions and soil-related corrosion. Flexible plastic pipe materials are more adaptable to changes in temperature, ground movement and corrosion. Other environmental influences include internal and external stresses, such as pipe pressure, traffic loading and bedding conditions that change local soil conditions. However, this information is not always available.

Operational. The effects of operational factors are difficult to determine since there is limited data. Nevertheless, some operational factors have been reported as necessary, such as pipe pressure, pumping, valve changes, flushing events or general operational activities. These operational factors can lead to high pipe failure through pipe fatigue and stress. The distance and time between pipe failures suggests some failures are related, potentially arising because of proximal changes to the environment from adjacent pipe repairs and consequent reduced resilience in pipes. This phenomenon is a form of cascading failure.

A combination of factors manifests concurrently to cause pipe failure, where environmental and operational conditions exert stresses on pipes with low resilience. The combination of factors is unique for each geographic region, material type and WDN. The literature reviewed and presented several studies investigating important factors, yet data remains inconsistent and incomplete across large networks. This review provides a foundation of information and a starting point to understand which factors may correlate with pipe failures in any given WDN.

In the wider context, the research undertaken provides a comprehensive foundation for any data scientist or industry professional developing pipe failure models or seeking to understand pipe failures.

Objective 2: Investigate historical data from a UK water company to understand the main causes of failure and determine suitable variables for modelling.

Chapters 3 and 4 characterise the local network and geographical conditions that influence pipe failure, using historical data from a large UK WDN. Various methods were used to understand operational variables, including spatial relationships, joints, large users, flushing events and water sources; pipe pressure data was unavailable. Dijkstra's path was used to calculate the shortest distance between failures, constrained by the pipe network. Pipe failures were observed to be spatially related, where several failures were observed within 20 m of another. Yet there was limited evidence to suggest these failures were related in time, with most failures occurring well more than a month of the first failure. Pipes carrying water sourced from reservoirs showed higher failure rates, particularly those of metal construction. Water from reservoirs is colder during the winter due to its exposure, which affects metal pipes due to the thermal conductive properties, in turn resulting in higher pipe failure during the winter months.

The environmental variables were investigated using a multivariate Generalised Additive Model (GAM) and Partial Dependence Plots (PDP), which were

demonstrated to provide a flexible way of revealing the complex inter-relationships and influences of environmental covariates. The results revealed a higher failure frequency for iron pipes during the winter. This is a consequence of the peat and clay soils retaining higher volumes of rainwater from the autumn months, which in turn freeze and cause thermal stress and ground movement as the soil expands. AC pipes have a higher failure frequency during the summer, where high temperatures and low rainfall can lead to high SMD and associated effects of highly expandable clay soils. PVC pipes also have a higher failure frequency during the summer, partly influenced by soil movement, but also other factors, such as pipe pressure and water usage, are more likely to cause issues; however, data on these factors were unavailable. Both SDI and PE were more resistant to pipe failure, but prolonged periods of frost were shown to marginally increase the failure frequency. Iron and AC exhibit stronger failure rate patterns with the variables due to the number of failures collected in the historical data. Since WDNs are increasingly adopting more resilient materials, like PE and SDI, there should be a greater emphasis placed on understanding the influences on failure rates for these materials, currently less understood.

In the wider context, some of the factors (for example, large users, flushing events, long periods of low temperature or high SMD) investigated here have not previously been reported, and are important for data scientists or industry professional to further understand the causes of pipe failures. The use of alternative methodologies such as the GAM model and PDPs provides an effective and easy analysis of the data that can be used in future research.

Objective 3: Compose and evaluate historical statistical pipe failure models to understand a potentially successful modelling approach.

Pipe failure data and the complexities of pipe failures makes modelling pipe failures inherently problematic. Numerous advances have been made over the previous decades, from simple single variate approaches to complex machine learning methods. Evaluating the various pipe failure models and their performance is essential to gaining an understanding of the most reliable model

and current gaps in thinking. In the targeted review, the evolution of statistical models was explored by composing and evaluating important literature on deterministic, probabilistic and machine learning methods developed to predict pipe failures.

The review concluded that the response of the model, prediction interval and the spatial level is influenced by time-dynamic variables such as weather, where seasonal effects can only be captured on short intervals (weekly, monthly, or quarterly). Models predicting short intervals are achievable, but it would require grouping the data at a very high spatial level (grouping pipes or network) to reduce the data imbalance, and at this scale, from a practical decision-making point of view, the results may not be as useful. Use of the results on such a spatial scale. Improved accuracies can be achieved by increasing the prediction interval to periods longer than annual. However, this limits the use of the model to long-term asset planning and abrogates the use of time-dynamic variables, which are important for describing some pipe failures observed in the network.

In the wider context, the review is useful for data scientists or industry professional who wish to gain an understanding of the various statistical models that can be used to predict pipe failures. More importantly to industry professionals the review provides an important overview of the different response types, prediction intervals spatial levels and levels of inference. These are important when choosing a model, therefore the review provides a decision aid to help data scientists determine the most appropriate approach, given different decision-making contexts.

Objective 4: Develop a pipe failure modelling methodology that identifies pipe failures and the risk of failure and apply it to the UK WDN.

Based on the information gained through the evaluation of statistical models in Objective 3, a pipe failure model was developed using a gradient boosting decision tree. In a practical context, the accuracy here provides the greatest utility when applied in conjunction with expert knowledge to inform decision makers.

Better accuracy may be achieved if predicting over five-year AMP periods. As with most machine learning methods, the gradient boosting tree model could not accommodate left truncated data (being data absent due to short reporting periods), which meant that previous unrecorded failures remained unaccounted for. This inevitably affects the results, but the full consequences of this were not exhaustively established and so this is an area meriting further investigation. Boosting is also prone to overfitting, but here the hyper-parameters were tuned using cross-validation techniques to ensure the best results. However, the tuning effort was extensive and the resultant model computationally expensive and time consuming.

The pipes were grouped by similar characteristics to allow more accurate results. Grouping is typically undertaken at a network-wide scale, yet this does not represent local conditions influencing pipe failures or fails accurately capture the historical failure rates of pipe groups. To reduce the effects of these limitations, pipes were grouped on a 1 km local scale, by similar characteristics. However, grouping in this way can alter the probability of failure for some pipes that cross boundaries. For example, the boundary may separate a single pipe into two separate 1 km grid, or soil zone, with one half of the pipe containing the failure and the other no failure. This would result in a lower-than-expected probability of failure for the pipe with no failure. Furthermore, the approach used may not be as useful for rural areas where fewer pipes are present, where smaller scales may be more appropriate (e.g., 1:100,000 is a smaller scale than 1:100). To accommodate this, future research could focus on the use of fuzzy logic to create the boundaries of the segmentation within a range.

The variables identified as being necessary from the investigations of Objective 2 were utilised, and regularisation was used as a form of variable selection, an advantage of machine learning methods. The most significant variables included number of previous failures, pipe length, SDM > 120 mm, pipe diameter and pipe age. Data on pipe pressure (e.g., transient or cyclical) would have been valuable as this has previously been correlated with pipe failures (Farmani et al., 2017;

Farrow et al., 2017), especially in those failures observed in PVC pipes during the summer. Other useful variables could include local bedding conditions, external pipe loading (e.g., traffic), pipe protection and conveyed water quality. Over-generalised local conditions influenced the model accuracy, and in this study, the local soil conditions were presented derived from a soil map at 1:250,000 scale. Likewise, the weather variables were highly generalised to an annual scale from a 40 x 40 km grid source. Additional problems were inherent to the network data, including the timing and location of pipe failures, which was dealt with to some extent in pre-processing, by using annual prediction intervals and then grouping pipes within localised areas.

Many pipe failure model studies do not address the critical step of understanding the risk pipe failure but stop at predicting the probability of failure. The approach here extends existing risk analysis approaches by weighting the consequences, thereby placing more importance on consequences of greater import. This is key for supporting management decisions that are often hindered by budget restraints and are limited to only replacing or rehabilitating typically small areas of the network. In this study, a weighted risk analysis was used, calculating the combination failure probability and the sum of the consequences, in this case, only water loss (pipe diameter) and potential damage (proximity to nearest property) were available. The results presented a metric that was useful for categorising risk cartographically when using Jenks natural breaks (Figure 6-6).

In the wider context, this methodology provides a reliable model that is provides a similar accuracy to comparable models and is useful to water companies to approximate the future behaviour of pipes. This may help to proactively manage pipe networks and help to meet leakage targets.

Objective 5: Discuss and synthesise the challenges and limitations of modelling pipe failures with professionals currently in practice.

Several scale and model iterations were used to determine the best approach for this research, yet overall, a poor accuracy was revealed to arise from many

approaches, particularly when trying to predict individual pipes or for shorter prediction intervals. These issues have also been reported in several different models in the literature, which often highlight data quality and quantity as a distinct barrier (Kabir et al., 2016; Mailhot et al., 2000; Tang et al., 2019). Several uncertainties and challenges arise when developing pipe failure models. To understand these data issues, a series of semi-structured interviews were organised with experienced industry practitioners to characterise current practices in the sector. Supported by literature and further discussions on pre-processing and model scalability, the study therefore provides a holistic view for data scientists and water professionals.

The establishment of smart networks (being currently a technology at an inception stage in the UK (Water Industry Journal, 2018)) will provide reliable and consistent real-time data. This will start to reveal local physical mechanisms of pipe failure that will potentially improve the accuracy of pipe failure models relying less on failure data captured in the field. However, this data alone cannot capture all the variables, and in many studies important variables are absent from data sets. More variables correlated with pipe failures must be collected such as pipe pressure and lining and coating of pipes, which was a significant limitation within this study. Additional variables are also needed, such as those investigated in this research, including local bedding conditions, excavation damage and network maintenance. There is a need for these, amongst others to be collected. There is also a wider discussion on management issues and company structures that inhibit data sharing. Collecting more information may be achievable by promoting data sharing and improving data collection by establishing collection protocols and data pools. However, achieving accurate short-term predictions is likely be challenging to achieve unless vast historical datasets are collected, since such predictions are mathematically problematic to model due to the temporal infrequency of pipe failures. For this reason, operational teams are unlikely to adopt the use of pipe failure models since they rely on short-term failure predictions. However, long-term decision making can still use failure models for asset management decisions, and, with data-driven approaches, improved data

collection and quality, and the use of machine learning, model accuracy is likely to improve outcomes.

In the wider context, understanding the fundamental issues with data quality and some of the limitations to modelling is important to both data scientists and industry professionals. This is important to start a conversation on how best to improve the accuracy of pipe failure models through improving data.

8.3 Implications of the Proposed Modelling Approach

Water companies seek innovative means to reduce pipe failures and are currently implementing statistical models to help achieve this. This study is therefore not only timely but offers a key aid in proactive management and in addressing Ofwat targets set to reduce leakage. This research provides an understanding of pipe failures, and a methodology that can benefit water companies.

8.3.1 Understanding Pipe Failures

There are several implications arising from this research in understanding and predicting pipe failures. The research enables decision-makers through revealing causation factors and relationships between several factors, revealing those most important for the UK WDN. These findings are transferrable to other UK WDNS and other countries with similar environmental and operational conditions. The findings can be used by data scientists to improve pipe failure models, or at an engineering level can support taking constructive actions towards building future resilience and extending the life of pipes.

However, further dedicated data collection is required to improve this understanding further. Particularly focussing on the quantity and quality of data and introducing more data from other factors such as operational maintenance, localised bedding conditions, traffic loading, water usage or operational stress amongst others. Furthermore, much is understood about materials such as Iron and AC which are older and more abundant materials, and therefore have richer associated datasets and more failures to correlate with data. Yet these materials are being phased out now in favour of relatively new plastics and SDI, which in

this UK case study, had attracted less data because there are to-date fewer pipes. As more of this material is used, more data will be collected in future and the reasons for failures will start to emerge. For this reason, a global failure model is more likely to provide better accuracy, since it can accumulate more pipe failures in the model and so improve predictive accuracy.

8.3.2 Using Advanced Methods (Machine Learning)

Gradient boosting trees offer an advanced method that has shown the best results in many studies and has proved useful in this methodology. One benefit of machine learning is the ability to determine complex relationships with large multifaceted data sets, which is advantageous if real time continuous data is collected in the future, e.g., in 'smart' networks. Another advantage of the method is the use of regularisation, which removes the need for variable selection methods, such as backward or forward selection or stepwise selection. The gradient boosting tree method also removes the need to standardise the data since the approach is not sensitive to the magnitude of input variables. It is often reported that machine learning is difficult to interpret; however, the gradient boosting tree model adopted in this methodology is classed as grey-box supervised learning approach, and benefits from interpretable variable importance and model accuracy measures. Ultimately the gradient boosting tree methodology can produce better accuracy if longer predictions are used and is more accurate than simpler methods such as Generalised Liner Models (GLM) or Generalized Additive Models (GAM) (see Appendix D.2). Better predictions will be useful for asset planning, although traditional statical approaches are still useful since they are quick, familiar, and easy to build, less computationally expensive and do not require hyper parameter tuning.

8.3.3 Determining Risk for Decision Makers

The methodology provides a means to guide replacements proportionate to the potential risk through a weighted risk assessment. A weighted risk assessment was suitable; however, several important data in such an approach remain unavailable. Further research into risk analysis in conjunction with a suitable pipe

failure model is considered likely to yield the most useful results. Outputs from such an approach could be interpreted by almost anyone and visualised in a GIS platform for simple inspection as well as being used for supporting proactive management decisions.

8.3.4 Reducing Leakage

The methodology could contribute to aiding management decisions and proactively managing WDNs by predicting pipes that are at a high risk of future failure. This will help work towards Ofwat's requirements to reduce leakage, in turn helping to meet the increasing water demands in the future. However, pipe failure models are unlikely to resolve the issue of future demand independently. Alternative means to reducing water, such as lowering water consumption and operational means to reduce real time for water loss, are needed in parallel with pipe failure models to meet necessary targets.

9 Conclusion

Reducing water loss, increasing serviceability, and meeting the commitment to reduce leakage has encouraged water companies to manage pipe failures proactively. One key area in proactive management is the development of statistical modelling approaches for predicting pipe failures and the risk of failure. Adoption of statistical models, particularly machine learning has reflected the increased availability and cost-performance of computational speed and data resource capacities over the past few decades, including many data-driven techniques now available to help understand and predict pipe failures. Adopting such approaches offers opportunities to manage assets proactively by aiding decision-makers with useful asset information.

A range of factors being potentially correlated with pipe failures were investigated. These were separated into five common material types to help understand the unique and complex relationships. The results obtained are transferable for other temperate climate countries, and for other UK networks that seek to understand pipe failures. As discussed however, important variables were not always available. This is a position often reported in pipe failure modelling studies. Collecting and investigating extended variable data could help to characterise and explain pipe failure further. Considering this, available failure records are also short, and the collection of continuous data sets over longer periods is needed in future. The findings do however enable support for decision-makers through revealing causation and exposing relationships between different the factors that were available, revealing those most important variables for the UK WDN needed for building a parsimonious and statistically significant pipe failure model.

A review of the literature on existing pipe failure models was undertaken to aid an understanding as to which approach might prove the most reliable. The findings indicated that several models can be used interchangeably, yet it is important to carefully consider available variables, the desired prediction interval, the spatial level and the response type, before selection of the most appropriate approach. Although complex modelling approaches such as machine learning

can show marginal improvements in predictive accuracy, it may perhaps be suggested that a slight decline in predictive accuracy be acceptable in exchange for an easier, potentially faster, and more interpretable model, and one that organisations can easily implement.

A gradient boosting tree model and weighted risk analysis methodology was developed. A single global model showed preferred results over individual material-specific models since some of the materials had too few pipe failure records for appropriate convergence; SDI and PE being relatively new pipe materials and having consequently fewer historic records. The most relevant variables discovered from Objectives 1 and 2 were used in the model, and regularization was used as a suitable selection method, being a particular advantage of machine learning methods. This shift towards machine learning has been justified by its reported improved prediction accuracy over traditional methods, and the potential to provide innovative ways to proactively manage the WDN. Often the probability of failure alone is not considered enough to determine the need for a pipe's replacement, and so assessment of the risk of failure was achieved through a weighted risk analysis. Although the results provided an acceptable accuracy, when used in conjunction with expert 'domain' knowledge, the model provides a useful approximation and offers a means to help decision makers plan rehabilitation and replacement efforts based on a reproducible and pragmatic approach. The accuracy could be improved by increasing the prediction interval to five-year AMP periods, which would also benefit longer term planning, and by including pipe pressure data. Further useful research would include estimating pipe failures for DMAs. This is an area that is particularly under-represented in literature, and one that has been highlighted as useful to water companies. Several challenges and limitations were revealed during the modelling process, primarily pertaining to the quality and quantity of the data and the difficulty of modelling imbalanced data, which prompts suggestions for future research.

Evaluating the challenges and limitations of the pipe failure model methodology, a series of semi-structured interviews were undertaken with experienced sector practitioners. The results revealed several issues which could be addressed to improve future pipe failure data. These discussions highlighted how many variables are unavailable or are not routinely collected, and how some data is siloed, with budget constraints from regulatory requirements inhibiting relevant data collection. These issues cause many datasets to be incomplete due to an absence of essential variables, and available data is often summarised and not reflective of true localised conditions. Other issues were identified when collecting data on-site, including a lack of understanding and time constraints, resulting in incomplete or inaccurate data, which leads to further pre-processing issues, such as trying to relocate pipe failures to the nearest pipe. Considering the scope and constraints of available data, improved data collection would be beneficial within water companies. Furthermore, with advances in smart networks, and real-time data collection, future opportunities may arise to collect useful data from smart sensors.

Insight gained into how sector professionals view the methodology presented here suggests that the approach would indeed be useful in the industry. In particular, the methods trialled can aid proactive long-term asset management in WDNs, transforming the decision-making process. This research also highlights a means of reducing subjectivity in decision-making through the application of data-driven information. However, operational teams having a more operational focus are unlikely to find this approach useful and are likely to find more useful approaches in real-time monitoring using smart sensors. This methodology provides a suitable approach for water companies to develop a reproducible and pragmatic model to predict pipe failures and the risk of failure. Such research is timely since there is a drive to reduce water lost through pipe failures and help water companies to achieve KPI objectives and reduce the chance of heavy regulator fines. Statistical models can aid with long-term proactive asset management, and lead to greater WDN resilience, operational reliability, and

reduced leakage, ultimately helping with the broader needs of future water demands.

9.1 Further Research Recommendations

Attempting to make informed, proactive maintenance decisions, especially determining the time for the first inspection, maintenance intervention, or replacement, is hindered by uncertainties in the data, pre-processing steps, spatiotemporal scaling, and model development. Based on the findings of the research, there are several recommendations that can be put forward to advance this work further:

- **Promote the collection of more variables related to pipe failures.**

The inclusion of more complex operational and network management variables is considered likely to result in better predictive accuracy for the model. In this study, several variables were used, yet these were constrained by availability. Opportunities to increase the number of variables is detailed in Chapter 7, and this can provide a starting point for future work.

- **Integrating data sources from real-time data collection.**

Smart sensors and the Internet of things offers ubiquitous, high resolution, real-time data. Smart sensor data may help improve pipe failure models by enabling the collection of pipe data and proximal soil characteristics on a finer temporal and spatial granularity than currently available. Yet, there is a challenge in this to manage the resultant volume of data and to use it to satisfy the needs of the water company.

- **Data engineering.**

Techniques such as feature extraction, deep feature synthesis, feature space analysis or data fusion techniques could offer a promising way to process variables, potentially improving models. Limited efforts have been made in this area for pipe failure modelling, but it provides an avenue of potential interest.

- **Rescaling pipe lengths.**

An alternative means of segmenting pipes (rather than simply grouping them) might improve pipe failure models. Some studies have used pipe length based on road sections (Winkler et al., 2018), or the average length of replacement (De Oliveira et al., 2011a).

- **Pipe size-specific models.**

There is a suggestion that models perform better when pipe cohorts are divided into different sizes, with some focussing on large diameter pipes (Rajeev et al., 2014; Wilson et al., 2017). This distinction was also discussed with industry professionals who suggested this was something that will benefit the results and is currently used in some instances. It would be worth investigating to establish if dividing small and large pipe diameters into separate models could yield more accurate results.

- **Alternative modelling scenarios.**

Predicting counts of failures is useful for risk-based models, and a desire to predict counts by DMA has been discussed. Predicting counts of failures is mathematically challenging. However, using an approach that recognises the inconsistency of failure intensity due to seasonal variations and changing management approaches would be suitable. The non-homogeneous Poisson process allows for this and introduces prior distributions that include expert opinion, which may help in the absence of essential variables (Tang, 2018).

- **Improve weighted risk assessment by including more consequences of failure.**

Understanding the risk of failure constitutes a critical element in asset management. The risk of failure should include a wide range of consequences, yet only a few such consequences were available in this approach. Expected consequences include loss of water, potential disruption, water quality, reliability, direct costs (damage to property and infrastructure and pipe repair and replacement) and indirect costs (environmental and social) (Kakoudakis et al., 2018). Future studies should seek to incorporate such metrics in the modelling.

- **Future implementation.**

The model developed has been tested on historical data from a UK case study, yet it has not been implemented into use. A logical next step would be to use the model in practice and to provide feedback on the results and their utility for supporting decision-making.

- **Include climate change conditions.**

Further work should also include projections of future climate scenarios from the UKCP18. These data can be used to explore scenarios based on longer-term impacts of extreme weathers, such as long dry periods, that may lead to more severe soil shrink-swell or prolonged periods of frost.

9.2 References

- Aslani, B., Mohebbi, S., Axthelm, H., 2021. Predictive analytics for water main breaks using spatiotemporal data. *Urban Water J.* 00, 1–16. <https://doi.org/10.1080/1573062X.2021.1893363>
- Barton, N.A., Hallett, S.H., Jude, S.R., 2021. The challenges of predicting pipe failures in clean water networks: a view from current practice. *Water Supply*. <https://doi.org/10.2166/ws.2021.255>
- Barton, N.A., Hallett, S.H., Jude, S.R., Tran, T.H., 2022. An evolution of statistical pipe failure models for drinking water networks: a targeted review. *Water Supply*. <https://doi.org/10.2166/ws.2022.019>
- Chen, T.Y.J., Guikema, S.D., 2020. Prediction of water main failures with the spatial clustering of breaks. *Reliab. Eng. Syst. Saf.* 203, 107108. <https://doi.org/10.1016/j.res.2020.107108>
- De Oliveira, D.P., Garrett, J.H., Soibelman, L., 2011. A density-based spatial clustering approach for defining local indicators of drinking water distribution pipe breakage. *Adv. Eng. Informatics* 25, 380–389. <https://doi.org/10.1016/j.aei.2010.09.001>

- Farmani, R., Kakoudakis, K., Behzadian, K., Butler, D., 2017. Pipe Failure Prediction in Water Distribution Systems Considering Static and Dynamic Factors. *Procedia Eng.* 186, 117–126. <https://doi.org/10.1016/j.proeng.2017.03.217>
- Farrow, J., Jesson, D., Mulheron, M., Nensi, T., Smith, P., 2017. Achieving zero leakage by 2050: the basic mechanisms of bursts and leakage, UK Water Industry Research Limited. London.
- Jenkins, J.M., Kowalski, M., Alvarenga, E.F.S., 2018. Predictive modelling of water losses using random forests on weather covariates. *Water Supply* 18, 2180–2187. <https://doi.org/10.2166/ws.2018.044>
- Kabir, G., Tesfamariam, S., Loeppky, J., Sadiq, R., 2016. Predicting water main failures: A Bayesian model updating approach. *Knowledge-Based Syst.* 110, 144–156. <https://doi.org/10.1016/j.knosys.2016.07.024>
- Kakoudakis, K., Farmani, R., Butler, D., 2018. Pipeline failure prediction in water distribution networks using weather conditions as explanatory factors. *J. Hydroinformatics* 20, 1191–1200. <https://doi.org/10.2166/hydro.2018.152>
- Levinas, D., Perelman, G., Ostfeld, A., 2021. Water Leak Localization Using High-Resolution Pressure Sensors. *Water* 13, 591. <https://doi.org/10.3390/w13050591>
- Mailhot, A., Pelletier, G., Noël, J.-F., Villeneuve, J.-P., 2000. Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application. *Water Resour. Res.* 36, 3053–3062. <https://doi.org/10.1029/2000WR900185>
- Rajeev, P., Kodikara, J., Robert, D., Zeman, P., Rajani, B., 2014. Factors contributing to large diameter water pipe failure as evident from failure inspection, in: *LESAM 2013 - IWA Leading-Edge Strategic Asset Management*. Sydney, Australia, pp. 9–14.

- 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. *Reliab. Eng. Syst. Saf.* 196, 106754. <https://doi.org/10.1016/j.ress.2019.106754>
- Tang, K., 2018. Statistical Modelling Approaches to Improve Risk Assessments for United Kingdom's Water Infrastructure. Cranf. Univ. Cranfield University.
- Tang, K., Parsons, D.J., Jude, S., 2019. Comparison of automatic and guided learning for Bayesian networks to analyse pipe failures in the water distribution system. *Reliab. Eng. Syst. Saf.* 186, 24–36. <https://doi.org/10.1016/j.ress.2019.02.001>
- Water Industry Journal, 2018. Reducing leaks: the rise of leak detection technology in water pipes [WWW Document]. URL <https://www.waterindustryjournal.co.uk/reducing-leaks-the-rise-of-leak-detection-technology-in-water-pipes> (accessed 10.7.20).
- Wilson, D., Fillion, Y., Moore, I., 2017. State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water J.* 14, 173–184. <https://doi.org/10.1080/1573062X.2015.1080848>
- Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., Tscheikner-Gratl, F., 2018. Pipe failure modelling for water distribution networks using boosted decision trees. *Struct. Infrastruct. Eng.* 14, 1402–1411. <https://doi.org/10.1080/15732479.2018.1443145>
- Yamijala, S., Guikema, S.D., Brumbelow, K., 2009. Statistical models for the analysis of water distribution system pipe break data. *Reliab. Eng. Syst. Saf.* 94, 282–293. <https://doi.org/10.1016/j.ress.2008.03.011>

Appendices

Appendix A Supplementary Material for Chapter 2

A.1 Supplementary Methods

The following shows the R Markdown code used to create the failure rate graphs in the manuscript using a bi-variate approach. Pipe failures recorded in time f was divided by pipe operating length l divided accordingly for each parameter class z_i , and used to estimate a standardised failure rate y (expressed as total bursts/1000 km/year as per Ofwat Reporting guidance on mains repairs (Ofwat, 2018) the relevant observation period T as $y = \frac{f/l}{T}$. By assuming a Poisson distribution (Asnaashari et al., 2013; Wols et al., 2019) of the data the variance $V(X) = \sigma^2$ is estimated.

A.2 R Script – Failure Frequency

The R script for pipe failure frequency is available from the following GitHub repository: <https://github.com/nabarton/Proactively-managing-drinking-water-distribution-networks-A-data-driven-statistical-modelling-app>

A.3 References

- Asnaashari, A., McBean, E.A., Gharabaghi, B., Tutt, D., 2013. Forecasting watermain failure using artificial neural network modelling. Can. Water Resour. J. 38, 24–33. <https://doi.org/10.1080/07011784.2013.774153>
- Ofwat, 2018. Final reporting guidance for PR19 2 Exclusions.
- Wols, B.A., Vogelaar, A., Moerman, A., Raterman, B., 2019. Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. Water Supply 19, 404–416. <https://doi.org/10.2166/ws.2018.085>

Appendix B - Supplementary Material for Chapter 3

Figure B-1 shows the results for distance between failures for up to 1 km of each other.

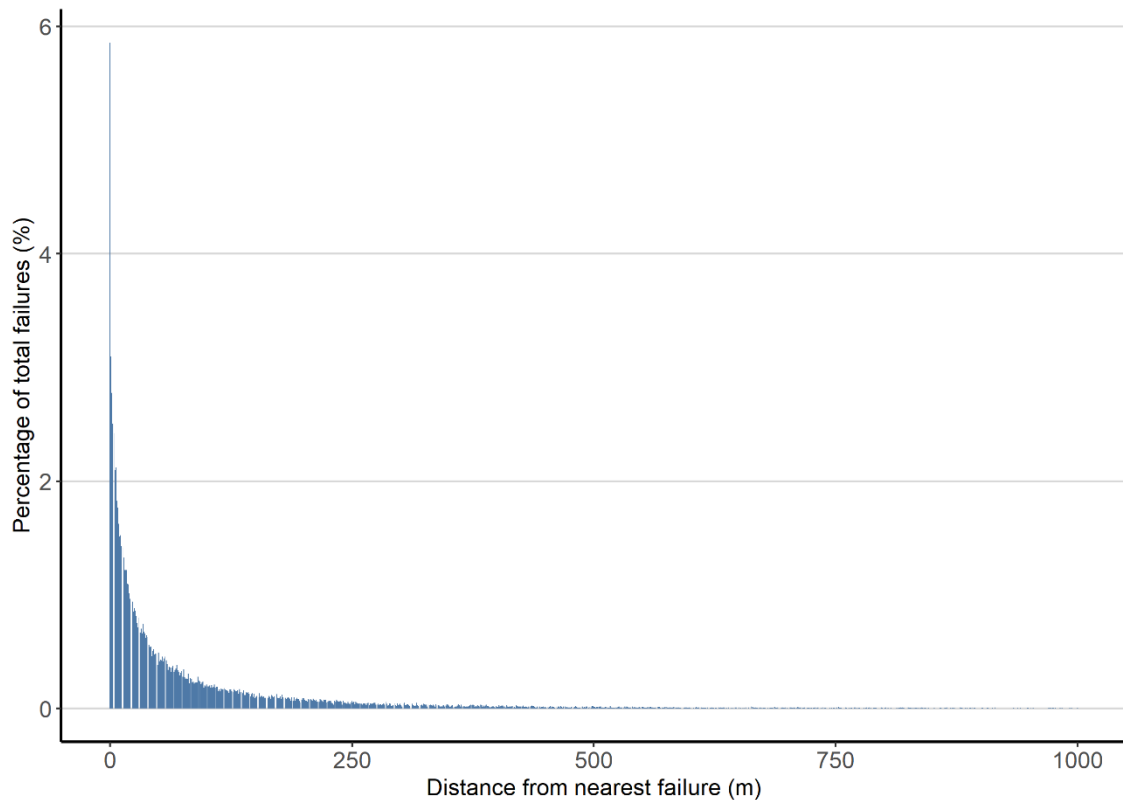


Figure B-1 Spatial analysis of all pipe failures within 1 km of each other. Data from UK water company for the operational period from 2005 to 2018.

Appendix C Supplementary Material for Chapter 4

C.1 Supplementary Variable Information

Table C-1 details the shrink swell classifications and Table C-2 details the hydrology of soil categories.

Table C-1 Shrink swell potential categories 1 – 6.

SSWELL Class	Shrink swell potential	Shrinkage (vol) %	Soil characteristics
Very low	1	< 3	Lithoskeletal, gravelly, sandy, light loamy or silty
Low	2	3 - 5	Medium loamy or medium silty unless derived from clay-with-flints/plateau drift
Moderate	3	5 - 12	Medium loamy or medium silty and derived from clay-with-flints or clayey on soft shales, Reddish Marl, clay and sand, non-swelling clay, or any till or head deposits
High	4	12 -15	Clayey and on swelling clay, clay-with flints/plateau drift or glaciolacustrine clay
Very High	5	> 15	Clayey and on brownish swelling clay
High	6	12 - 15*	Alluvial clay or peat – very high SSWELL. *Potential that is not realised unless effective drainage is installed to at least 2m depth.

Table C-2 Hydrology of Soil Classes based on substrate hydrogeology (presence of aquifers, catchment and groundwater relationships, soil type and approximate depth to aquifer) (Boorman et al., 1995).

1 = Weakly consolidated, microporous, by-pass flow uncommon.	1 – Soft sandstone, weakly consolidated sand. 2- Weathered/fissured intrusive/metamorphic rock.
2 = Weakly consolidated, macroporous, by-pass flow uncommon.	3- Chalk, chalk rubble.
3= Strongly consolidated, non or slightly porous, by-pass flow common.	4- Soft magnesium, brashy or Oolitic limestone and ironstone. 13- Hard (fissured) sandstones. 14- Earthy Peat. 15- River alluvium.
4= Unconsolidated, macroporous by-pass flow very uncommon.	5- Hard fissured limestone. 7-Hard but deeply shattered rocks.
5= Unconsolidated, macroporous by-pass flow common.	6- Hard coherent rocks. 8- Soft shales with subordinate mudstones and siltstones. 9- Very soft reddish blocky mudstones (marls). 10- Very soft massive clays. 11- Very soft bedded loams, clays and sands. 12- Very soft bedded loam/clays/sand with subordinate.
6= Slowly Permeable.	16- Marine alluvium. 18- Colluvium. 21- Glaciolacustrine clays and silts. 24- Gravel. 26- Chalky drift.
7= Impermeable (hard).	17- Lake marl or tufa. 19- Blown sand. 22- Chalky drift. 27- Disturbed ground.
8= Impermeable (soft)	20- Cover loam. 23- Clay with flints or plateau drift. 25- Loamy drift.
9= Eroded Peat.	28- Eroded Blanket Peat.
10= Raw Peat.	29- Raw Peat.

C.2 Supplementary Results

The material specific generalized additive model summary tables are shown in Table C-3 to Table C-7.

Table C-7Table C-3 Generalized additive model summary for Iron pipes.

Parametric Terms:	df	Chi.sq	p-value	
Diameter_Band	3	3125.78	< 0.00	
Age_Band	5	511.26	< 0.00	
Shrink_Swell	4	96.55	< 0.00	
Subsoil	4	295.09	< 0.00	
HOST	6	623.5	< 0.00	
Approximate significance of smooth terms:	edf	Ref.df	Chi.sq	p-value
s(Temp)	8.18	9	1186.81	< 0.00
s(TchgW1)	4.45	9	450.24	< 0.00
s(TchgW2)	8.29	9	152.74	< 0.00
s(DaysAirFrost)	6.99	9	521.61	< 0.00
s(SMD)	7.27	9	579.03	< 0.00
s(SMDchgW1)	8.17	9	131.6	< 0.00
s(SMDchgW2)	3.94	9	15.37	< 0.00
s(pH)	7.59	9	274.5	< 0.00

Table C-4 Generalized additive model summary for Asbestos Cement pipes.

Parametric Terms:	df	Chi.sq	p-value	
Diameter_Band	3	1035.97	< 0.00	
Age_Band	2	857.79	< 0.00	
Shrink_Swell	4	74.24	< 0.00	
subsoil	4	82.67	< 0.00	
HOST	6	92	< 0.00	
Approximate significance of smooth terms:	edf	Ref.df	Chi.sq	p-value
s(Temp)	7.96	9	84.61	< 0.00
s(TchgW1)	1.15	9	2.79	0.09
s(TchgW2)	7.78	9	51.69	< 0.00
s(DaysAirFrost)	5.39	9	78.18	< 0.00
s(SMD)	6.24	9	589.39	< 0.00
s(SMDchgW1)	1.72	9	24.33	< 0.00
s(SMDchgW2)	2.09	9	18.12	< 0.00
s(pH)	8.83	9	353.91	< 0.00

Table C-5 Generalized additive model summary for Steel & Ductile Iron pipes.

Parametric Terms	df	Chi.sq	p-value	
Diameter_Band	3	151.13	< 0.00	
Age_Band	2	30.12	< 0.00	
Shrink_Swell	4	54.73	< 0.00	
Subsoil	4	33.56	< 0.00	
HOST	6	103.13	< 0.00	
Approximate significance of smooth terms:	edf	Ref.df	Chi.sq	p-value
s(Temp)	1.10	9	1.77	0.15
s(TchgW1)	2.15	9	11.24	< 0.00
s(TchgW2)	2.45	9	9.21	< 0.00
s(DaysAirFrost)	2.76	9	23.69	< 0.00
s(SMD)	1.56	9	3.97	0.06
s(SMDchgW1)	2.15	9	4.79	0.05
s(SMDchgW2)	0.48	9	0.29	0.46
s(pH)	5.55	9	48.92	< 0.00

Table C-6 Generalized additive model summary for Polyvinyl Chloride pipes.

Parametric Terms:	df	Chi.sq	p-value	
Diameter_Band	3	41.3	< 0.00	
Age_Band	3	43.45	< 0.00	
Shrink_Swell	4	19.25	< 0.00	
Subsoil	4	19.07	< 0.00	
HOST	6	34.98	< 0.00	
Approximate significance of smooth terms:	edf	Ref.df	Chi.sq	p-value
s(Temp)	5.55	9	87.03	< 0.00
s(TchgW1)	0.78	9	1.71	< 0.00
s(TchgW2)	7.16	9	35.16	< 0.00
s(DaysAirFrost)	3.082	9	21.19	< 0.00
s(SMD)	3.35	9	31.24	< 0.00
s(SMDchgW1)	4.01	9	14.39	< 0.00
s(SMDchgW2)	0.54	9	0.66	0.26
s(pH)	5.76	9	59.95	< 0.00

Table C-7 Generalized additive model summary for Polyethylene pipes.

Parametric Terms:	df	Chi.sq	p-value	
Diameter_Band	3	13.10	< 0.00	
Age_Band	2	63.08	< 0.00	
Shrink_Swell	4	7.33	< 0.00	
Subsoil	4	22.11	< 0.00	
HOST	6	36.98	< 0.00	
Approximate significance of smooth terms:	edf	Ref.df	Chi.sq	p-value
s(Temp)	3.48	9	40.08	< 0.00
s(TchgW1)	0.53	9	1.31	0.09
s(TchgW2)	2.16	9	4.79	0.06
s(DaysAirFrost)	4.48	9	41.45	< 0.00
s(SMD)	0.67	9	0.47	0.41
s(SMDchgW1)	1.18	9	2.49	0.10
s(SMDchgW2)	2.65	9	8.59	0.01
s(pH)	4.41	9	13.29	0.01

C.3 R Script – Generalised Additive Model and Partial Plots

The R script for the generalized additive model is available from the following GitHub repository: <https://github.com/nabarton/Proactively-managing-drinking-water-distribution-networks-A-data-driven-statistical-modelling-app>

C.4 References

Boorman, D.B., Hollis, J.M., Lilly, A., 1995. Hydrology of soil types: a hydrologically based classification of the soils of United Kingdom, Institute of Hydrology.

Fasiolo, M., Nedellec, R., Goude, Y., Wood, S.N., 2019. Scalable Visualization Methods for Modern Generalized Additive Models. *J. Comput. Graph. Stat.* 1–9. <https://doi.org/10.1080/10618600.2019.1629942>

Appendix D Supplementary Material for Chapter 6

D.1 Supplementary Variable Information

Table D-1 presents the soil shrink swell categories, Table D-2 presents the hydrology of soil categories and Table D-3 presents the corrosion to iron categories.

Table D-1 Shrink swell potential categories 1 – 6.

SSWELL Class	Shrink swell potential	Shrinkage (vol) %	Soil characteristics
Very low	1	< 3	Lithoskeletal, gravelly, sandy, light loamy or silty
Low	2	3 - 5	Medium loamy or medium silty unless derived from clay-with-flints/plateau drift
Moderate	3	5 - 12	Medium loamy or medium silty and derived from clay-with-flints or clayey on soft shales, Reddish Marl, clay and sand, non-swelling clay, or any till or head deposits
High	4	12 - 15	Clayey and on swelling clay, clay-with flints/plateau drift or glaciolacustrine clay
Very High	5	> 15	Clayey and on brownish swelling clay
High	6	12 - 15*	Alluvial clay or peat – very high SSWELL. *Potential that is not realised unless effective drainage is installed to at least 2m depth.

Table D-2 Hydrology of Soil Classes based on substrate hydrogeology (presence of aquifers, catchment and groundwater relationships, soil type and approximate depth to aquifer).

1 = Weakly consolidated, microporous, by-pass flow uncommon.	1 - Soft sandstone, weakly consolidated sand. 2- Weathered/fissured intrusive/metamorphic rock.
2 = Weakly consolidated, macroporous, by-pass flow uncommon.	3 - Chalk, chalk rubble.
3= Strongly consolidated, non or slightly porous, by-pass flow common.	4 - Soft magnesium, brashy or Oolitic limestone and ironstone. 13 - Hard (fissured) sandstones. 14 - Earthy Peat. 15 - River alluvium.
4= Unconsolidated, macroporous by-pass flow very uncommon.	5 - Hard fissured limestone. 7-Hard but deeply shattered rocks.
5= Unconsolidated, macroporous by-pass flow common.	6 - Hard coherent rocks. 8- Soft shales with subordinate mudstones and siltstones. 9- Very soft reddish blocky mudstones (marls). 10- Very soft massive clays. 11- Very soft bedded loams, clays and sands. 12- Very soft bedded loam/clays/sand with subordinate.
6= Slowly Permeable.	16 - Marine alluvium. 18- Colluvium. 21- Glaciolacustrine clays and silts. 24- Gravel. 26- Chalky drift.
7= Impermeable (hard).	17 - Lake marl or tufa. 19- Blown sand. 22- Chalky drift. 27 - Disturbed ground.
8= Impermeable (soft)	20 - Cover loam. 23- Clay with flints or plateau drift. 25- Loamy drift.
9= Eroded Peat.	28 - Eroded Blanket Peat.
10= Raw Peat.	29 - Raw Peat.

Table D-3 Corrosion to iron categories and description.

Corrosivity Class	Description
1	Non-aggressive
2	Slightly aggressive
3	Moderately aggressive
4	Highly aggressive
5	Very highly aggressive

Table D-4 Natural Soils Resource Institute (NSRI) clay vulnerability data set - Dominant clay vulnerability class between 1 – 9.

Clay Class	Description
1	Extremely low
2	Very low
3	High
4	Low
5	Medium low
6	Medium
7	Medium High
8	Very High
9	Extremely High

D.2 Alternative Models

The complexity of pipe failures should require advanced modelling methods. However, studies have shown that both traditional and data-driven methods can perform well, especially in the presence of data limitations (Alizadeh et al., 2019; Chen et al., 2019). Four different models were developed, a simple parametric Generalised Linear Model (GLM), a semi-parametric Generalised Additive Model (GAM) model, not often explored in pipe failure modelling, and two data-driven non-parametric tree models, a random forest and gradient boosting model, both have demonstrated better predictions when compared to other data-driven

methods (Giraldo-González and Rodríguez, 2020; Snider and McBean, 2019; Winkler et al., 2018; Wols et al., 2019).

Generalized Linear Model

GLMs are discussed in Section 5.4.1.3. In GLM models, a differentiable monotonic link function describes the relationship between the covariates and the dependent variable, allowing for the use of non-normal distributions, in this instance the logit distribution: $g(p) = \beta_0 + \sum_{j=1}^p \beta_j x_j + \varepsilon_i$, where (p) is the probability of failure, $g(\cdot)$ the logit distribution $\log\left(\frac{P(X)}{1-P(X)}\right)$, where P is the probability and X the n -dimensional input dataset (see Section 5.4), and β_j is the location term of some predictor covariate x_j (Dunn and Smyth, 2018).

Regularization was used to penalise the coefficients with low variance towards zero, simplifying the model and reducing overfitting (Heinze et al., 2018). In this scenario, the elastic net was used to lead to better predictions whilst eliminating variables with no correlation (Hastie et al., 2009). The shrinkage rate λ determines how the model penalises insignificant features, here set between 0.01 and 0.1. These complexity parameters were returned through cross-validation. The GLM was developed using the ‘glmnet’ package from Rs CRAN repository (Friedman et al., 2021).

Generalized Additive Model

The GAM extends the GLM to include a smoothing term which makes the model semi-parametric. In this respect it is more flexible than the GLM. The full model description, including smoothing technique and regularisation is presented in Section 4.2.2. However, in this case, like the GLM model, the logit distribution is used to predict the probability of pipe failure. The GAM model was developed using the ‘mgcv’ package from Rs CRAN repository (Wood, 2019).

Random Forest

The random forest classification model is an ensemble model based on the bagging technique. A detailed model description of a classification-based tree

mode, bagging technique and regularisation is presented in Section 5.4.3.5. The random Forest model was developed using the ‘caret’ package from Rs CRAN repository (Kuhn et al., 2020)

Gradient Boosting Trees

Gradient boosting tree classification model is based on the boosting ensemble method. A full description of the approach is presented Section 0.

Results

The accuracy measure results of the different models are presented in Table D-5 showing the Balanced Accuracy, Matthews Correlation Coefficient (MCC), Area Under the Curve (AUC) and Brier’s Score. The results show the gradient boosting tree model is statistically more accurate, with a balanced accuracy of 0.65 and an MCC score of 0.27 which suggests the model is better than the alternatives at predicting both classes, although only marginally. The Brier score is the same for GBT, GAM and RF models, yet observing the calibration curve shows a better fit for the GBT model (Figure D-2). The AUC score is highest for both the GAM and the GBT models. Ultimately the GBT was found to provide the best fit, yet the GAM and RF are both acceptable at predicting accurate classes. The GLMNET performs poorly compared to the other models when considering the accuracy measures. This is likely a result of the parametric linear approach being unable to describe complex relationships.

Table D-5 Accuracy Evaluation Metrics for Test Dataset

Measure	GLMNET	GAM	RF	GBT
Balanced Accuracy	0.53	0.63	0.65	0.65
Matthew’s Correlation Coefficient	0.17	0.24	0.25	0.27
Area Under the Curve	0.86	0.89	0.88	0.89
Brier’s score	0.008	0.007	0.007	0.007

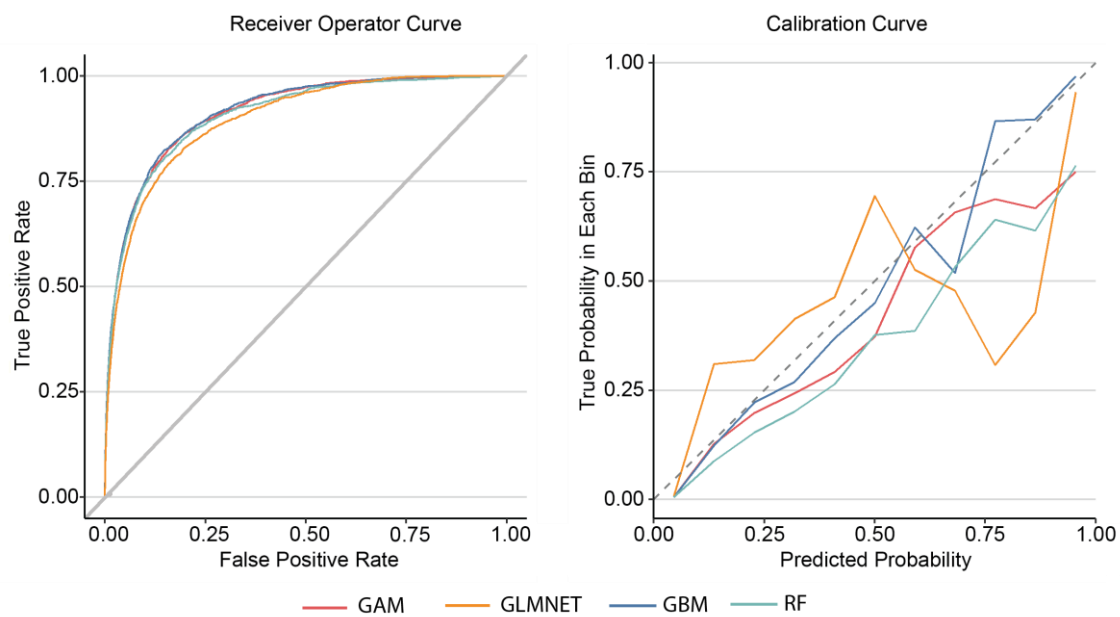


Figure D-2 ROC curves and calibration curves for the original data, the over sampling, and the weighted classes.

Conclusion

This analysis compared different modelling techniques for predicting the probability of pipe failure. The evaluation metrics for all models infer reasonable accuracy, however; when appraising the difference, the gradient boosting model shows a higher MCC value, lower Brier's score and a better fitted calibration curve. The limited variation between the models demonstrates only marginal improvements in predictions achieved using more advanced modelling techniques such as the gradient boosting model, nonetheless; these improvements may prove significant to the final decision-makers. Since the gradient boosting model has the better statistical accuracy, this model was adopted for use in the pipe failure methodology in Chapter 6.

D.3 Sampling techniques for imbalanced data

There has been discussion in pipe failure modelling literature on the importance of balancing data (Robles-Velasco et al., 2021). Alternatively others suggest that balancing may not always improve results (Barton et al., 2021; Kleiner and Rajani, 2012). Two different sampling techniques, class weighting, and ROSE

oversampling were compared to the original imbalanced data to determine the effects of balancing the data when predicting the probability of failure. Under sampling was excluded on the argument that too much information is lost when there is an extreme data imbalance. Likewise, common methods of oversampling were excluded since it duplicates the same minority class multiple times. ROSE over sampling however synthetically generates random minority samples that balances the classes to create a 50:50 data balance. The full details can be found in Lunardon et al. (2021). Finally, the class weighting imposes a heavier costing when errors are made on the minority class. The three approaches were applied to the modelling methodology presented in Chapter 6.

The accuracy measure results of the three methods have been presented in Table D-6, showing the Balanced Accuracy, MCC, AUC and Brier's Score. The results show that the Balanced Accuracy and MCC are higher in the original data set, suggesting that both classes (failures and non-failures) are predicted better in this scenario. This is also supported by the Balanced Accuracy.

The AUC, is higher for the class weighted and oversampling methods, suggesting they have a better discriminative ability to distinguish between the classes, yet the difference is marginal. This can be seen in Figure D-3. However, the Briers score suggests that the predictions in the original data has calibrated closer to those observed, which is observed further in Figure D-3 that clearly shows the difference between the three methods. This suggests that the sampling techniques do not calibrate well when predicting the probability of failure, as observed in other studies (Fan et al., 2022).

Table D-6 Accuracy Evaluation Metrics for Test Dataset

Measure	Original	Class weighted	Over sampling
Balanced Accuracy	0.65	0.47	0.40
Matthew's Correlation Coefficient	0.27	0.18	0.17
Area Under the Curve	0.89	0.91	0.91
Brier's score	0.007	0.097	0.098

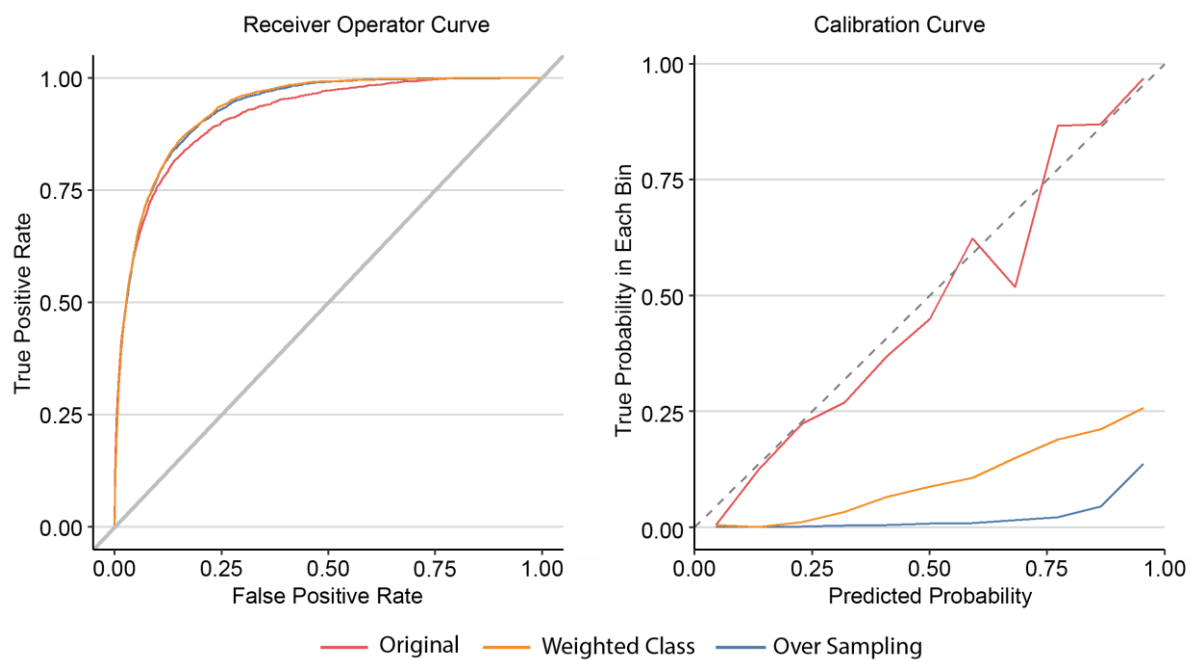


Figure D-3 ROC curves and calibration curves for the original data, the over sampling, and the weighted classes.

Based on the results, the pipe failure methodology developed in this research should use the original dataset without weighting or oversampling the data.

D.4 R Script – Gradient Boosting Decision Tree

The gradient boosting decision tree R script is available from the following GitHub repository: <https://github.com/nabarton/Proactively-managing-drinking-water-distribution-networks-A-data-driven-statistical-modelling-app>

D.5 References

- Alizadeh, Z., Yazdi, J., Mohammadiun, S., Hewage, K., Sadiq, R., 2019. Evaluation of data driven models for pipe burst prediction in urban water distribution systems. *Urban Water J.* 16, 136–145. <https://doi.org/10.1080/1573062X.2019.1637004>
- Barton, N.A., Hallett, S.H., Jude, S.R., 2021. The challenges of predicting pipe failures in clean water networks: a view from current practice. *Water Supply*. <https://doi.org/10.2166/ws.2021.255>
- Chen, T.Y.-J., Beekman, J.A., David Guikema, S., Shashaani, S., 2019. Statistical Modeling in Absence of System Specific Data: Exploratory Empirical Analysis for Prediction of Water Main Breaks. *J. Infrastruct. Syst.* 25, 04019009. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000482](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000482)
- Dunn, P.K., Smyth, G.K., 2018. Chapter 10: Models for Counts: Poisson and Negative Binomial GLMs. Springer, New York, NY, pp. 371–424. https://doi.org/10.1007/978-1-4419-0118-7_10
- Fan, X., Wang, X., Zhang, X., ASCE Xiong (Bill) Yu, P.E.F., 2022. Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors. *Reliab. Eng. Syst. Saf.* 219, 108185. <https://doi.org/10.1016/j.ress.2021.108185>
- Friedman, J., Hastie, T., Tibshirani, R., Narasimhan, B., 2021. Package “glmnet.”
- Giraldo-González, M.M., Rodríguez, J.P., 2020. Comparison of Statistical and Machine Learning Models for Pipe Failure Modeling in Water Distribution Networks. *Water* 12, 1153. <https://doi.org/10.3390/w12041153>
- Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics, Springer Series in Statistics.

- Springer New York, New York, NY. <https://doi.org/10.1007/978-0-387-84858-7>
- Heinze, G., Wallisch, C., Dunkler, D., 2018. Variable selection – A review and recommendations for the practicing statistician. *Biometrical J.* 60, 431–449. <https://doi.org/10.1002/bimj.201700067>
- Kleiner, Y., Rajani, B., 2012. Comparison of four models to rank failure likelihood of individual pipes. *J. Hydroinformatics* 14, 659–681. <https://doi.org/10.2166/hydro.2011.029>
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Ziem, A., Scrucca, L., Hunt, T., 2020. Package ‘caret’ - Classification and Regression Training.
- Lunardon, N., Menardi, G., Torelli, N., 2021. Package ‘ROSE.’
- Robles-Velasco, A., Muñuzuri, J., Onieva, L., Rodríguez Palero, M., 2021. Trends and applications of machine learning in water supply networks management. *J. Ind. Eng. Manag.* 14, 45. <https://doi.org/10.3926/jiem.3280>
- Snider, B., McBean, E.A., 2019. Improving Urban Water Security through Pipe-Break Prediction Models: Machine Learning or Survival Analysis. *J. Environ. Eng.* 146, 04019129. [https://doi.org/10.1061/\(ASCE\)EE.1943-7870.0001657](https://doi.org/10.1061/(ASCE)EE.1943-7870.0001657)
- Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W., Tscheikner-Gratl, F., 2018. Pipe failure modelling for water distribution networks using boosted decision trees. *Struct. Infrastruct. Eng.* 14, 1402–1411. <https://doi.org/10.1080/15732479.2018.1443145>
- Wols, B.A., Vogelaar, A., Moerman, A., Raterman, B., 2019. Effects of weather conditions on drinking water distribution pipe failures in the Netherlands. *Water Supply* 19, 404–416. <https://doi.org/10.2166/ws.2018.085>

Wood, S.N., 2019. Package “mgcv” - Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.

Appendix E Supplementary Material for Chapter 7

E.1 Ethical Approval Letter



10 January 2021

Dear Mr Neal Barton,

Reference: CURES/12565/2021

Title: Predicting Pipe Failures using Forecast Weather Data

Thank you for your application to the Cranfield University Research Ethics System (CURES).

We are pleased to inform you your CURES application, reference CURES/12565/2021, has been approved. You may now proceed with the research activities you have sought approval for.

If this approval is for research conducted as part of your studies, please remember to include a copy of this letter with your thesis submission.

OR

If this approval is for research conducted as part of a course of study, please remember to include a copy of this letter with your thesis submission.

If you have any queries, please contact CURES Support.

We wish you every success with your research.

Regards,

CURES Team

E.2 Questionnaire

The following presents the general questions and alternative questions used for the semi-structured interview are presented in Table E-1.

Table E-1 Interview Topic guide

Interview code:	Date:
Introduction	
<p>Welcome to the interview session and thank you for deciding to take part in this activity as part of the project 'Predicting pipe failures using forecast weather data'.</p> <p>This research project is funded by the National environmental Research Council in collaboration with Cranfield University. This research aims to support water companies in future by helping to reduce network leakage.</p> <p>Participation on this activity is completely voluntary and you are free to withdrawn at any time. If you agree, the interview session will be recorded to fully capture the flow of information from participant to researcher. The extent of your participation would be limited to an interview, which should not be longer than 60 minutes.</p> <p>This research will comply with the rules specified by the UK Data Protection Act. All the information given will be kept strictly confidential and your name will be substituted with an anonymous participant code in all records kept. The only person that will have access to records of the sessions and recordings is the researcher and these will be stored securely for the duration of the research and then destroyed. If you have any concerns about the conduct of this study, you may contact researchdata@cranfield.ac.uk</p> <p>The PhD student Neal Barton neal.a.barton@cranfield.ac.uk is currently engaged in the project and is supervised by Steve Hallett s.Hallett@cranfield.ac.uk Simon Jude s.jude@cranfield.ac.uk at Cranfield University.</p> <p>Do you have any questions about the research project?</p> <p>Do you have any questions about this activity?</p> <p>Would you like to participate?</p>	
The interviewee	
<p>Job role:</p> <p>What is your current role?</p> <p>What is the scope of your work?</p> <p>Which department of the business is your role allocated to?</p> <p>Responsibilities:</p>	

What are your main responsibilities?
How are your responsibilities delivered?

Methodology

Main Questions	Alternative Questions (If needed)
Do you think this model is useful?	How would you see this model applied within your organisation?
What is your opinion on the model scale?	Strategic/operational/long term/short term?
What are your thoughts on the variables used?	Should any additional variables be included/excluded? What is your opinion on the variable importance?
What is your opinion of model accuracy?	What level of accuracy would you like to see? Do you think it is important?

General Discussion

Main Questions	Alternative Questions (If needed)
How would you typically decide which pipes are going to be replaced, rehabilitated, or repaired?	Do you consider the future failure rates of pipes? How are these determined?
What sort of information do you obtain to help inform this decision?	What factors are considered? Which factors are the most important?
What is your understanding of pipe failure models and their applicability?	Do you use pipe failure models within the organisation? In your opinion, what are the pros and cons?
What is your understanding of machine learning?	Do you have any examples in your organisation?
What type of data do you collect to inform pipe failures?	On site - how is the data collected/recorded? Network data - explain how network data is maintained/updated? Historical data? What are the notable challenges when collecting pipe failure data?

In your opinion, are any important data unavailable or not routinely collected?	Why are these unavailable?
Do you use smart sensors to detect leakage or for pipe failures?	Can you give any examples? What are the advantages and disadvantages?
In your opinion, what drives data collection?	Could you expand on this, and provide an example?
How accessible is data within your organisation?	Is there a central database or form of integrated data storage? In your opinion what are the barriers and enablers to integrated data sharing? Have you experienced data silos?
What are the main organisational challenges to sharing data?	What do you consider the enablers and inhibitors of data transfer? Are you and your team aware of what other parts of the business do, and what data they collect? How is knowledge of data availability promoted by the organisation?
How is data access controlled in your business?	Any data challenges such as GDPR, data rights and data security etc? How does this limit data collection? How does user agreements limit your ability to work with data?
How would you describe the quality of data collected and shared?	Can you give any examples? Are there any other notable issues? Are there any notably good examples?
Can you think of any ways you would promote data sharing and transparency within your organisation?	What could be done differently?
Are there any successful data sharing examples that you can provide?	Any dedicated internal resources to help managers?
Are you aware of big data technologies for data control, and are any of these currently used in the business?	

Does the organisation have necessary expertise for big data technologies?

Does the business support training?
Do you outsource work to experts?

Are you aware of data ontologies?

Can you give an example within your organisation?

Would you see the benefit in establishing a data ontology?

Is this something you think could be easily implemented into your organisation?

What do you think the enablers would be?
What do you think the inhibitors might be?

Wrap up

Would you like to add anything else?

Are there any other topics you would like to further explore in the context of the project?
