Imperial College London
Department of Civil and Environmental Engineering

Impact of pressure fluctuations on pipe failures in water distribution networks

Hossein Rezaei

July 2017

Submitted in part fulfilment of the requirements for the degree of
Doctor of Engineering in Water Engineering of Imperial College London
and the Diploma of Imperial College London

CONFIDENTIAL

This thesis is confidential and will be embargoed for two years as per standard Imperial College policy for theses containing commercial work.

During the period of the embargo, please do not distribute except under the terms of the STREAM Non-Disclosure Agreement.

To Rumi, the poet of love.

Abstract

Water utilities operate ageing infrastructures that are degraded by environmental factors and operational stresses. Pipe failures have become a routine, resulting in major interruptions and extensive costs to the society.

Pipe failure is a result of complex interactions between a variety of factors contributing to the pipe's structural degradation and ultimate failure. Previous studies have extensively reviewed environmental and pipe-related factors. However, long term impact of quasi-steady and unsteady pressure variations on metallic pipe failures is not fully investigated.

The overall aim of this project is to enhance the understanding of the dynamic pressure variations in water supply networks and evaluate their impact on pipe failures, with motivation to enhance the operational efficiency of water supply infrastructures by managing systems' hydraulic conditions.

In this study, a large-scale survey sampling programme is designed and executed in order to gather network representative high frequency pressure samples. A metric is formulated to quantify the stresses imposed on pipes from pressure variations. Causal analysis is undertaken and relationship between pipe failure and predictor variables are investigated by developing logistic regression models. The study develops a methodology for investigating cost-effectiveness of intervention measures and economic justification of the calm networks.

The findings from the study illustrate positive associations between the system's hydraulic variations and predicted probability of pipe failure. It is shown that deterioration models can be enhanced by including pressure variation characteristics as contributing factors to pipe degradation. Investment in achieving calm networks is demonstrated to be economically justifiable.

Acknowledgements

I would like to express my appreciation and thanks to my academic supervisor, Dr. Ivan Stoianov, for his continuous support in development and delivery of this work. His vision, enthusiasm, energy and technical contributions have been hey to this project. I would also like to thank professor Nigel Graham, my other academic supervisor, for his strategic guidance and support, helping me to stay focused and think outside the box.

Special thanks to my industrial supervisor, Bernadette Ryan, for her committment to this project, excellent support, positive attitude and encouragement, when it was needed the most. I would like to also thank all the staff at Severn Trent Water who supported this project over the years, in particular, the research and development team that welcomed me warmly and helped me fining my way around the company. I would like to thank all people who helped with the data collection programme, in particular Sally Rumens for her technical advice, and Pat Hesketh for tirelessly playing a key role in delivery of the field work.

I would like to acknowledge EPSRC and Severn Trent Water for funding this project. Special thanks to the Stream Engineering Doctorate Centre for creating an enjoyable experience in undertaking this project.

I would like to thank all staff at Imperial College London that supported this project, Dr. Michael Templeton for his tutoring and all fellow friends and colleagues at InfraSense Labs, specially Hooman Armand, for countless brainstorming and simulating discussions.

I would like to express my gratitude to my family, for all their love and support. Above all, I would like to express my greatest thanks to my wife, Roya, for her continuous love and encouragement over the years.

Declaration

I herewith certify that all material in this dissertation which is not my own work has been properly acknowledged.

The copyright of this thesis rests with the author and is made available, once the embargo expires, under a Creative Commons Attribution-Non Commercial-No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or distribution, researchers must make clear to others the licence terms of this work.

Abbreviations

AC	Asbestos Cement
AIC	Akaike Information Criterion
AMP6	Asset Management Period (period 6)
ANOVA	Analysis of variance
ARR	Accounting Rate of Return
ASTM	American Society for Testing and Material
AUROC	Area under the ROC curve
AWWA	American Water Works Association
AZNP	Average Zone Night Pressure
BADC	British Atmospheric Data Centre
BCR	Benefit-Cost Ratio
Capex	Capital Expenditure
CBA	Cost-Benefit Analysis
CEA	Cost-Effectiveness Analysis
CG	Control Group
CI	Cast Iron
CIPRA	Cast Iron Pipe Research Association
CPIS	Cumulative Pressure Induced Stress
CS1	Case Study (area) 1
CS2	Case Study (area) 2
CS3	Case Study (area) 3
DI	Ductile Iron
DM	Deterioration Modelling
DMA	District Metered Area
DOF	Degree of Freedom
Dojm	An enterprise database system at STW
DP	Dynamic Pressure
EPSRC	Engineering and Physical Sciences Research Council
FCV	Flow Control Vales

Continued on the next page

FP	(soil) Fracture Potential
FT	Filed Test
GIS	Geographic Information Systems
GLM	Generlized Linear Model
GUI	Graphical User Interface
KPI	Key Performance Indicator
LL	Log Likelihood
mAOD	Metres Above Ordnance Datum
MC	Monte-Carlo
MCA	Multi Criteria Analysis
MoE	Margin of Error
Netbase	A Water Network Management System
NPV	Net Present Value
ODI	Outcome Delivery Incentive
Ofwat	Office for Water
OR	Odds Ratio
PE	Polyethylene
PRV	Pressure Reducing Valve
PT	Pressure Transient
PV	Pressure Variation
PVC	Polyvinyl Chloride
ROC	Receiving Operating Characteristic
ROCOF	Rate of Occurrence of Failure
ROI	Return on Investment
SAP	Systemanalyse und Programmentwicklung-Enterprise
SCADA	Supervisory Control And Data Acquisition
SCC	Stress Corrosion Cracking
SD	Standard Deviation
SDB	Supply Demand Balance
SE	Standard Error
SIM	Service Incentive Mechanism
SSLRC	Soil Survey and Land Research Centre
STREAM	Industrial doctoral centre for the water sector
STW	Severn Trent Water
SW	Sample Weighting
Totex	Total Expenditure

Continued on the next page

TRL	Technology Readiness Level
UKWIR	UK Water Industry Research
WDIN	Water Distribution Infrastructure Network

Symbols

β_0	Logistic regression model coefficient
$\delta\sigma$	Nominal applied stress (fluctuating over a range)
δK_{eff}	Effective stress intensity parameter
δK	Amplitude of the range of alternating crack-tip stress
ϵ	Strain
$\frac{da}{dN}$	increment of crack extension per load cycle
ρ	Density of water
$\sigma_{\max}$	Maximum stress
$\sigma_{\min}$	Minimum stress
σ_{a_e}	Fatigue strength
σ_a	Stress amplitude
σ_f	Failure stress
σ_h	Standard deviation of stratum h
σ_i	Amplitude of pressure variations
σ_m	Mean stress
σ_r	Range of stress
a	Crack size
A	Area
c_h	Direct cost of individual sample from stratum h
c	Wave celerity or propagation speed
CF	Cash flow
d_a/d_N	Propagation of fatigue crack
D	Pipeline diameter
f	Darcy-weisbach friction factor
g	Gravity acceleration
H	Piezometric head
i	Discount factor
K_c	Fracture toughness (critical stress intensity)
K_f	Fatigue notch factor

Continued on the next page

K_T	Theoretical stress concentration factor
K	Bulk modulus
LL	Log-likelihood
M	Number of different cyclic loading events
m	Number of parameters in the model
n_h	Sample size for stratum h
N_h	Population size for stratum h
n_i	Number of applied stress (or strain) cycles
N_i	Fatigue life at a combination of stress amplitude and mean stress level
N_s	Number of strata
N_y	Number of years
P_{mean}	Diurnal mean pressure
P	Pressure
q	Notch sensitivity factor
Q	Flow discharge
R_f	Frictional resistance
R	Stress ration
S_c, N_c	S-N value in presence of corrosion/cracks
S_f	Fracture stress
S_u	Ultimate strength
S_y	Yield strength
S	Stress amplitude
SC	Fractional contribution of source
SW	Sample adjustment weighting
U	Closure ratio
V	Flow velocity
$W_{failure}$	Critical damage value
W_i	Product of stress and its number of cycles

Contents

1. Introduction **16**

1.1. Motivation and objectives 17

1.1.1. Problem statement 18

1.1.2. Gaps in the research area 19

1.1.3. Objectives 20

1.1.4. Scope and limitations 21

1.2. Novelty and contributions 22

1.3. Thesis structure 25

1.3.1. Keywords 26

2. Literature review **28**

2.1. Introduction 28

2.2. Mains and failure 29

2.3. Failure mechanisms and causes 32

2.3.1. Loads on mains 33

2.3.2. Corrosion 35

2.3.3. Joint failure 38

2.3.4. Ground movement 38

2.3.5. Temperature effect 39

2.3.6. Pipe age and size 40

2.3.7. Other factors 40

2.4. Failure modes 41

2.4.1. Gaps and limitations 45

2.5. Pipe deterioration modelling 46

2.5.1. Deterioration models 46

2.6. Pressure transient 53

2.7. Current state of the UK water industry 58

2.7.1. Data availability and quality 59

2.8. Summary 61

3. Methodology **63**

3.1. Introduction 63

3.2. Overview of research development 63
3.2.1. Study areas 64
3.2.2. Model case studies 64
3.3. Data 65
3.3.1. Missing data 68
3.4. Experimental programme 69
3.4.1. Field work procedure 69
3.4.2. Equipments and instrumentations 70
3.4.3. Burst data collection 71
3.4.4. Data analysis 73
3.5. Statistical analysis 73
3.6. Cost-benefit framework 73

4. Single-variate analysis of historical pipe failures 76
4.1. Company network analysis 76
4.1.1. Network characteristics-CS3 76
4.1.2. Single-variate analysis in CS3 82
4.2. Case study analysis 93
4.2.1. Network characteristics-CS1 93
4.2.2. Single-variate analysis in CS1 97
4.3. Summary and conclusion 100

5. Fatigue failure of water mains 101
5.1. Fatigue failure 101
5.1.1. Fatigue mechanism 103
5.1.2. Fatigue analysis and characteristics 107
5.2. Cumulative Pressure-Induced Stress (CPIS) 114
5.2.1. Cumulative stress 115
5.3. Summary 117

6. Experimental programme 118
6.1. High resolution pressure data sampling 118
6.1.1. Data collection and sampling 119
6.1.2. Stratified random sampling 120
6.1.3. Survey sampling bias and errors 120
6.1.4. Sample size 121
6.1.5. Sample size calculation 122
6.2. Stratified random sampling 123
6.2.1. Stratified random sampling - stage 1 123
6.2.2. Stratified random sampling - stage 2 131

6.3. Logging duration 131
6.4. Execution of the experiment 133
6.5. Verification of network representation 135
6.6. Summary 138

7. Pressure data analysis and CPIS development 139
7.1. Categories of pressure profiles 139
7.1.1. Network assets and pressure fluctuations 140
7.2. Event extraction and cycle counting 143
7.2.1. Inflexion coding 144
7.2.2. Cycle counting 145
7.3. CPIS across the network 150
7.4. Hydraulic tracing 151
7.4.1. Tracing techniques 152
7.5. Energy dissipation of hydraulic instabilities 155
7.5.1. Limitations 158
7.5.2. Simplified methods 158
7.5.3. Flow weighting method 159
7.5.4. Flow weighting method plus static headloss 159
7.5.5. Energy dissipation linear model 160

8. Multi-variate analysis of pipe failures 164
8.1. Multiple logistic regression 164
8.1.1. Odds ratio 167
8.2. Model statistical tests 167
8.2.1. Statistical significance 167
8.2.2. Goodness-of-fit 168
8.2.3. Validation of predicted values 170
8.3. Model case studies and datasets 171
8.4. Model development 172
8.4.1. Development of model 1 174
8.5. Final models 176
8.6. Summary 177

9. Model results and discussions 178
9.1. Statistical model results discussion 178
9.1.1. Models statistics 178
9.1.2. Estimated coefficients and the odds ratios 180
9.2. Effect plot 186
9.3. Summary 197

10. Cost-benefit framework for calm networks **198**
10.1. Background . . . 199
10.2. Costs of pipe failure . . . 200
10.2.1. A case study - United States . . . 200
10.2.2. Cost-benefit framework . . . 201
10.3. Methods of cost benefit evaluation . . . 202
10.3.1. Benefit cost ratio . . . 203
10.3.2. Net Present Value . . . 204
10.4. Methodology for cost-benefit framework . . . 205
10.4.1. Estimating benefits of pipe failure reduction and costs of prevention measures . . . 205
10.4.2. Distribution and range of cost and benefit estimates . . . 206
10.4.3. NPV, MC and sensitivity analysis . . . 207
10.4.4. Case study data . . . 208
10.5. Benefit of pipe failure mitigation . . . 208
10.5.1. Failure's direct cost estimation . . . 208
10.6. Cost of intervention activities . . . 217
10.7. NPV and MC simulation . . . 219
10.7.1. Sensitivity Analysis . . . 220
10.8. Research business case and outcome exploitation . . . 221
10.8.1. Research benefits . . . 222
10.8.2. Exploiting the outcome . . . 222
10.8.3. Knowledge transfer . . . 224
10.9. Chapter summary . . . 225

11. Conclusions and recommendations **226**
11.1. Summary of findings . . . 227
11.2. Novelty and key contributions . . . 228
11.3. Recommendations for future works . . . 228

References **230**

Appendices **243**
A. Experimental programme . . . 244
B. Statistical analysis . . . 249
C. Cost-benefit framework . . . 262

List of Tables

1.1. Research objectives, novelties and chapters where discussed . . 24

2.1. Pipe failure and the contributing factors 34
2.2. Failure modes and classifications-based on Vipulanandan et al. (2011) 41
2.3. Data requirements for various models 51
2.4. STW pipe with missing data in the Year 2008 60

3.1. Characteristics of the case studies 65
3.2. List of data attributes used in each analysis 66
3.3. Required data and general availability in STW 67
3.4. Data collected when repairing mains failure 74

4.1. Pipe failure factors considered for single-variate analysis 83
4.2. Combined soil characteristics versus historical burst distribution-metallic pipes 90

5.1. Comparison of SCC and fatigue corrosion-adopted from Roberge (1999) 110

6.1. Sample size for various confidence intervals and margins of error 122
6.2. Example of CG ranking based on CG pipe failure rate 125
6.3. Example of CG ranking based on their operation sites 126
6.4. Control group ranking based on their operational activities . . 126
6.5. Control group ranking based on their large industry consumers 127
6.6. Control group ranking based on type of mains burst 127
6.7. Example of cg ranking based on the performance of their DMAs 128
6.8. Combined rankings generating overall control group ranking . . 130
6.9. Characteristics of the Strata resulted from stage 1 131
6.10. Distribution of sampling points in each stratum 131

7.1. Example of DP calculation from rainflow cycle counting 149
7.2. Forward tracing supply contribution from logging sites 155
7.3. Factors for DP calculation using multi-variate linear model . . 162

8.1. Characteristics of the case studies used for model development 172
8.2. List of predictor variables considered for the models 173
8.3. List of models 174
8.4. Model coefficient estimates (Model 1) 175

9.1. Model statistics (Model 1) 179
9.2. Model coefficient estimates (Model 1) 181
9.3. Model's Odd Ratio and the 95% confidence intervals (Model 1) 181
9.4. Model statistics (Model 2) 183
9.5. Model coefficient estimates (Model 2) 183
9.6. Model's Odd Ratio and the 95% confidence intervals (Model 2) 183
9.7. Model statistics (Model 3) 184
9.8. Model coefficient estimates (Model 3) 184
9.9. Model's Odd Ratio and the 95% confidence intervals (Model 3) 184
9.10. Model statistics (Model 4) 185
9.11. Model coefficient estimates (Model 4) 185
9.12. Model's Odd Ratio and the 95% confidence intervals (Model 4) 185

10.1. Factors impacting total cost of pipe failure 209
10.2. Investment expenditure from the Company's business plan . . . 210
10.3. Annual cost of water lost from leakage 211
10.4. Direct cost of pipe repair 212
10.5. Characteristics of pipe cohorts 213
10.6. Property-minutes values for each cohort 214
10.7. Values used in sample size calculation 214
10.8. Sample size of pipe failure cohorts 214
10.9. Key costs of intervention measures 218
10.10Total failure reduction benefits and intervention costs 218
10.11Range of values used for calculating NPV 219
10.12Calculating present value for each year of 5-year plan 219
10.13Cost, benefit and NPV from MC simulations 220
10.14Sensitivity analysis for the 10 most important factors 221

11.1. Information and data to be collected for enhancing future analyses231

B.1. Soil shrink/swell level, adopted from Jarvis et al. (1997) 249
B.2. Soil corrosivity level, adopted from Jarvis et al. (1997) 250
B.3. Correlation matrix - Model 1 254
B.4. Correlation matrix - Model 3 254
B.5. Correlation matrix - Model 4 254

B.6. Correlation matrix - Model 2 255

List of Figures

1.1. Structure of the thesis with chapter numbers 26

2.1. Common joints used for CI and DI (a)Bell and spigot (b)Gland type (c)Tyton joint 30
2.2. Bathtub curve and shift in the wear-out phase due to additional stresses imposed on the pipe-adopted from Rajani and Tesfamariam (2004) 32
2.3. Conditions leading to corrosion - adopted from Rajani and Kleiner (2001) 35
2.4. Failure modes for (a) small and (b) large diameter mains-adopted from Makar et al. (2001) 42
2.5. Stresses imposed on pipe; longitudinal stress (left), radian stress (middle), hoop stress (right) 43
2.6. Deterioration model categories and example models(1) 47
2.7. Deterioration model categories and example models(2) 47
2.8. Other categorisation of deterioration models 49
2.9. Overview of the deterioration model at STW 52
2.10. Examples of data inaccuracy - distance of recorded burst location to the closest pipe and potential allocation to multiple pipes . . 60

3.1. Steps in collection of high frequency pressure data 70
3.2. InfraSense logging equipment, deployed on a fire hydrant 71
3.3. InfraSense platform GUI with plots of pressure data and monitoring site locations 72
3.4. Comparison of low (15 minutes interval) versus high (64Hz) resolution pressure data 72

4.1. Annual distribution of burst events 77
4.2. Distribution of pipe material and diameter 78
4.3. Distribution of pipe age with material and diameter 79
4.4. Distribution of pipes' diurnal mean pressure for metallic distribution ($\leq$ 250mm) and transmission ($>$ 250mm) pipes 80

4.5. Distribution of pipes' diurnal maximum pressure for metallic distribution ($\leq$ 250mm) and transmission ($>$ 250mm) pipes . . 81
4.6. Distribution of pipes' diurnal pressure variation for metallic distribution ($\leq$ 250mm) and transmission ($>$ 250mm) pipes . . 81
4.7. Distribution of metallic pipes' maximum pressure and pressure variation 82
4.8. Distribution of pipes with more than one historical bursts . . . 84
4.9. Distribution of burst events across mains material (a) number of bursts (b) normalised for mains length 85
4.10. Burst events distribution across mains diameters (metallic) (a) number of bursts (b) normalised for mains length 86
4.11. Distribution of mains burst events for mains age (metallic) . . 87
4.12. Cumulative monthly distribution of burst events in metallic pipes 88
4.13. (a)Burst distribution versus minimum daily temp over 7 days before burst (b)Burst distribution versus maximum daily temperature change within 7 days preceding burst date 89
4.14. Distribution of mains burst events versus soil fracture potential (a) number of bursts (b) normalised for mains length (metallic pipes) 91
4.15. Distribution of mains burst events versus soil corrosivity (a) number of bursts (b) normalised for mains length (metallic pipes) 92
4.16. CS1 (a) system schematic (b) pipe diameter size distribution . 95
4.17. Distribution of pipes' age and material in CS1 95
4.18. Distribution of pipes' age and diameter in CS1 (metallic pipes) 96
4.19. Distribution of pipes' material and diameter in CS1 96
4.20. Annual distribution of pipe failures in CS1 98
4.21. Cumulative monthly distribution of pipe failures in CS1 98
4.22. Distribution of pipe failures for mains age, normalised for mains length in CS1 98
4.23. Distribution of failures in pipe material (a) absolute (b) normalised for mains length 99
4.24. Distribution of failures in pipe diameter (a) absolute (b) normalised for mains length 99

5.1. Schematic of cyclic loading profile tension-tension cycle 102
5.2. An indicative unsteady pressure profile 102
5.3. Curve for total life, crack initiation and propagation 103
5.4. $S - N$ curve for different materials 104

5.5. Limits on critical pits to prevent crack growth with a stress of 90MPa-adopted from Marshall (2001) . . . 105
5.6. Crack growth rate development with accumulated cycles 106
5.7. Stress intensity factor vs. crack growth rate 107
5.8. Impact of mean stress on stress amplitude 108
5.9. Corrosion fatigue factors 109
5.10. Effect of corrosion on $S - N$ curve 111
5.11. An example for notch impact on fatigue life reduction-adopted from Davis (2004) 112
5.12. Impact of stress cycles on crack closure 113
5.13. Crack development caused by cyclic loadings on the pipe 114
5.14. Stress-strain curve for mild and brittle steel 115
5.15. Pipe failure due to pressure variations stress ranges 116

6.1. Stages of stratified random sampling 123
6.2. Locations of high frequency logging 135
6.3. Comparison of distribution for assets logged versus all pipes (entire Company) at (a) age groups (b) pipe diameter groups . 136
6.4. Comparison of distribution for assets logged versus all pipes (entire Company) at (a) soil corrosion groups (b) fracture potential groups 136
6.5. Comparison of distribution for assets logged versus all pipes (entire Company) at (a) material groups (b) pressure variation ranges 137
6.6. Comparison of distribution for assets logged versus pipes (entire Company) pipes at maximum pressure ranges 137

7.1. Steady-state profile with moderate pressure variations (Group 2) 141
7.2. Varying pressure profile with large pressure variation events (Group 3) 141
7.3. A large transient event on a trunk main as result of value operation (Group 4) 141
7.4. Impact of a PRV in mitigating a pressure transient event generated upstream (upstream profile is Group 4, downstream profile is Group 1) 142
7.5. Frequent and large pressure transient events (Group 5) 142
7.6. Impact of pump switches in generating pressure transient events (Group 6) 142
7.7. Data analysis process 144

7.8. Pressure database sizes using high and low resolution loggers . 145
7.9. Inflexion coding processing steps-from Hoskins (2015) 146
7.10. Comparison of raw and inflexion-coded data-from Hoskins (2015) 146
7.11. Cycle counting methods with cycles detected in each method, from Denk and Bakran (2013) 147
7.12. Rainflow counting algorithm by ASTM ASTM (2011) 148
7.13. Improved rainflow counting algorithm, from Hoskins (2015) . . 149
7.14. Distribution of DP values for sites logged across the network . 150
7.15. Distance between network nodes using (a) hydraulic tracing method, (b) Euclidean, (c) Manhattan 151
7.16. Development of index array-adopted from DNVGL (2014) . . . 154
7.17. (a) Source node (logging site FT501) (b) forward tracing analysis for FT501 154
7.18. Headloss when calculating DP at node A (*Method 2*) 160
7.19. Algorithm to estimate DP at each target node using *Method 2* 161
7.20. Example of calculating hydraulic shortest path and pipes characteristics (screen-capture from simulation run 162

8.1. Logistic probability curve 165

9.1. ROC curves for the models 180
9.2. Effect plot for age 187
9.3. Age group effect plot 188
9.4. Effect plot for diameter 189
9.5. Effect plot for length 191
9.6. Effect plot for material 192
9.7. Effect plot for Pmean 194
9.8. Effect plot for DP 196

10.1. Renewal decision for (a) low cost and (b) high cost failures-adopted from Rajani and Kleiner (2004) 202
10.2. Indicative relation of CPIS and other factors 204
10.3. Overview of methodology stages 206
10.4. Distribution of property.minute for interruption incidents in each cohort 215
10.5. Supply interruption costs of samples in each cohort (labels represent costs outside chart axis range) 216
10.6. Average supply interruption cost for each cohort per annum . . 216
10.7. Distribution of NPV values from MC simulation 220
10.8. Technology readiness stages-adopted from Mankins (2004) . . . 223

A.1. Pressure transient data collection form 245
A.2. Examples of corrosion and wall thickness loss 246
A.3. Stress concentration from external objects (tree roots) 247
A.4. Stress concentration from external objects (rocks) 247
A.5. Burst data collection tablet application 248

B.1. Kernel density of pipe segment length for various pipe material 251
B.2. AR CG 50m grid with mains, pipe failure records and customer connections 252
B.3. Distribution of customers distance to closest pipe 253
B.4. Correlation plots - Model 1 256
B.5. Correlation plots - Model 2 257
B.6. Correlation plots - Model 3 258
B.7. Effect plot for model 2 260
B.8. Effect plot for model 2-cont. 261

C.1. Pipe failure questionnaire - page 1 263
C.2. Pipe failure questionnaire - page 2 264

1. Introduction

Each year, there are tens of thousands of pipe failures across the water supply networks in the UK. In Severn Trent Water (STW) area alone, a network of approximately 46,000km pipes, supplying around 7.8 million customers, there are around 5000-7000 pipe failures per annum, only on the transmission and distribution mains (not including failures on connection and supply pipes). There are numerous direct and indirect costs associated with these failures, including the costs of repair, regulatory non-compliance fines, social and environmental costs, etc. resulting in millions of pounds damage.

Pipe failures in water supply network are often attributed to factors associated with pipe conditions (e.g. material, age), loadings (e.g. soil movement and traffic), environmental and external factors (e.g. climate impact) and their combined effect. In addition to these, intensive pipe cyclic pressure variations under steady and unsteady state hydraulic conditions contribute to structural degradation of pipelines, accelerating fatigue crack growth and ultimately a pipe failure. This research investigates the medium and long term impact of pressure variations resulting from the unsteady and quasi-steady state hydraulic conditions on fatigue related pipe failures.

Water utilities operate ageing infrastructures with many pipes reaching the end of their lifespan. Water supply systems are expected to meet an increasing demand due to the population increase and urbanisation. The current world population of 7.3 billion is expected to reach 8.5 billion by 2030, 9.7 billion in 2050 and 11.2 billion in 2100, according to the United Nations (2015). Today, around 54 per cent of the world's population lives in urban areas, a proportion that is expected to increase to 66 per cent by 2050. As water utilities are driven to pursue a sustainable and efficient operation, hydraulically calm networks could provide reduction in pipe failures and discolouration events and an extension of the lifespan of these critical assets. A hydraulically calm network is a network with minimum pressure variations resulted from unsteady and quasi-steady state hydraulic conditions.

Pressure-leakage relationships and adverse impacts of maximum pressure and Average Zone Night Pressure (AZNP) on mains deterioration have been extensively studied, e.g. Lambert (2001) and Lambert and Fantozzi (2010),

resulting in various improvement in the management of steady state pressure in water supply networks.

The primary focus in these studies have been on reducing average zone pressure. Little attention has been given to pressure variations and in particular, dynamic pressure variations. Water distribution systems are often assumed to operate under steady state conditions. This assumption is however frequently violated. Rapid changes in flow velocity cause significant pressure changes in the pipe, including pressure transients and quasi-steady state variations with different amplitude and cycle frequency.

Previous studies, such as Thorley (2004), have highlighted various failure incidents resulting from extreme pressure transients. These include the damage and deformation of pipe and joints, collapsing pipelines due to negative (sub-atmospheric) pressure and leakage into the pipes, pipeline vibration and movement, etc. While the impact of extreme pressure transients exceeding the pressure rating of a pipe are well studied, the medium and long term impact of the unsteady and quasi-steady state pressure variations on fatigue related cracks and deterioration in water supply systems have received little attention.

Unsteady pressure variations can result in pipe structural deterioration, through increased stress-induced corrosion rate, crack initiation, crack propagation and eventual fatigue failure. This is particularly important when combined with the diminishing tensile strength of pipes as a result of the development of micro cracks, corrosion and impacts from the environmental factors like temperature. Recent studies, such as Hoskins (2015) and Rezaei et al. (2015), highlighted the potential impact of pressure variations on pipe deterioration and failures.

1.1. Motivation and objectives

This research programme has been motivated by multiple operational challenges.

The initial drive for the research was to enhance the understanding of the unsteady-state pressure variations and their characteristics in a large scale operational network. In recent years, significant technological developments in instrumentation, data acquisition, data management and analytical techniques have enabled enabled the execution of complex experimental programmes to study this phenomenon. Of particular interest is enhancing the understanding of solid-fluid interactions, in particular the deterioration impact from cyclic pressure variations which reduce the lifespan of pipe assets. This research programme investigates the occurrence and long term impact of unsteady and

quasi-steady pressure variations in water supply networks on pipe failures.

The second driver was the derivation of performance indicators (metrics) which uniquely describe the unsteady and quasi-steady state pressure variations. These metrics are essential for improving the performance of pipe burst prediction models. Such pipe burst prediction models are increasingly being used for pipe rehabilitation and investment planning.

The insights into the dynamic hydraulic conditions of large scale operational networks which this research programme provides would also facilitate the implementation of Totex (Total expenditure) models, as introduced by the water regulator, Ofwat, in AMP6 (Asset Management Period, period 6 for year 2015-2020). Within Totex, operational expenditures and changes could result in more optimal performance; thus reducing the need for or deferring Capex (Capital expenditures).

Consequently, based on the results of this research programme and the concept of marginal gains, incremental changes and operational expenditures to minimise pressure variations can potentially lead to significant savings in capital expenditures for pipe rehabilitation and replacement, and burst repairs.

1.1.1. Problem statement

In response to continuous financial, regulatory and water resources pressures on water utilities to deliver efficiencies, UK water utilities have been pro-actively seeking control solutions for optimising pump scheduling and minimising average zone pressure (e.g. the proliferation of electronic controllers for flow modulation).

The increased level of control in operational networks could lead to increased hydraulic instabilities and unsteady pressure variations. There are many factors that cause rapid changes in the system's hydraulic characteristics resulting in the generation of unsteady and quasi-steady state pressure variations. These include: operational assets such as pump switches (planned or accidental), automated valve status changes, network interventions activities (including valve operation), pressure and leakage management activities, customers' behaviour and usage variation (especially industrial customers), etc.

Design of water mains generally includes a safety factor that takes into account unsteady pressure variations. However, in practice, such considerations are often limited. Hydraulic transient analysis of water supply systems is rarely carried out. Even where transient analysis is undertaken, surge protection schemes are not fully validated, mainly due to the lack of adequate monitoring equipment to acquire the necessary high-sampling rate and time-synchronised

data. While changes in the operation and control of water supply systems continuously occur, surge protection equipment is seldom re-validated and upgraded. Therefore, the hydraulic transient models might no longer be fully representative of the system, and surge protection devices not performing as expected. Additional factor is the absence of suitable maintenance of surge protection devices (e.g. air vessels, air valves, etc.) can lead to an unexpected and sub-optimal response when necessary.

Network physical and topological (connectivity) characteristics, leakage, friction, and other factors, detailed in section 2.6 chapter 2, dissipate the pressure waves. This, however, does not always lead to complete elimination of destructive energy. Various recent experimental studies, such as Hoskins (2015), and the data collected in this research work, highlight that unsteady pressure variations are frequently occurring across the water supply systems, often with noticeable amplitudes and frequencies.

1.1.2. Gaps in the research area

Medium to long-term exposure of pipe assets to unsteady and quasi-steady state pressure variations and their impact on pipe failures is little understood. Limited studies on fatigue impact of hydraulic loadings, such as Mohebbi et al. (2010), has been mainly limited to investigating structural deterioration of a particular pipe's characteristics (e.g. large diameter cast iron pipes) and the role of microstructure defects on fracture development. Often, these have focused on the material characteristics angle and usually only at the individual pipe level, rather than network view.

There have been numerous statistical approaches that have taken the pipe failure analysis beyond a single pipe level. These have considered a list of covariates that contribute to pipe failure and developed models that investigate correlations of the covariates and pipe failure. Majority of such models have been developed based on small sets of data (pipe failure records) and fail to consider the extend of covariate variabilities. These models and approaches are reviewed in detail in chapter 2. In larger networks, a significant shortcoming of these models is not considering factors related to the network's hydraulic performance and variations. The main reason for this is the gap in understanding the long term impact of hydraulic factors on fatigue-induced failures.

The lack of a systematic investigation of the occurrence of unsteady and quasi-steady state pressure variations and their impact on fatigue induced pipe failures in large scale operational networks is due to several factors. Firstly, it is the limited insights into the system dynamic behaviour. The unsteady hydraulic

modelling capabilities at network level are limited and existing modelling are often only practical for systems with a simple network topology. Traditionally, various simplifications were made when studying dynamic hydraulic phenomena, mainly due to the relative complexity of computer analysis tools, dealing with rather complex set of non-linear partial differential equations with sophisticated boundary conditions. Transportation and dissipation of the dynamics energy waves are often complex in large networks, specifically in networks with complex topologies.

Secondly, a significant barrier in development of knowledge in this area has been the lack of adequate instrumentation to gather and analyse operational high resolution dynamic data. Most water utilities do not have the technical tools and ability to continuously monitor the dynamic hydraulic performance of their systems. High resolution pressure data are needed in order to capture the detailed behaviour of the events. Conventional instruments and SCADA (Supervisory Control And Data Acquisition) systems, widely used across the water industry, do not have the necessary resolution and sensitivity to capture the rapid pressure fluctuations and transient events. The past decade has seen increasingly rapid advances in development of technology in data collection, instrumentation, data analytical tools and computation capabilities. There is still, however, a noticeable shortcoming in adopting and using the advanced technological tools in this area.

Thirdly, quality of pipe failure records is a limiting factor in large scale investigations. Pipe failure modes and mechanisms can assist in understanding pipe failure's root causes. However, for statistical investigation of pipe failures, limited and poor available records on historical failures events and the mechanisms can preclude development of detailed models. Historical records, including data on pipe failure events, are often prone to errors impacting the model reliability.

1.1.3. Objectives

The overall aim of this project is to carry out a systematic investigation of the occurrence of unsteady and quasi-steady state pressure variations and their impact on fatigue induced pipe failures in large scale operational networks. Specific objectives include:

I. Provide unique insight into the dynamic hydraulic behaviour across the water supply networks and study the unsteady pressure variations and their characteristics, types, amplitudes, frequencies, energy wave dissipation mechanisms, etc.

II. Review the existing knowledge on the factors affecting and accelerating pipe failure mechanisms and critically examine the existing deterioration modelling approaches and their shortcomings

III. Design and execute an experimental programme to gather system representative high frequency pressure data across a large supply network

IV. Systematically investigate associations between unsteady and quasi-steady pressure variations and pipe's deterioration and failure at the network level, using the advances in statistical techniques

V. Develop a cost-benefit framework to evaluate cost effectiveness of achieving calm networks. The framework is to consider economical, sustainable and practical factors and the thresholds at which intervention actions aiming calm networks are justified

1.1.4. Scope and limitations

The focus of this study is on *transmission* and *distribution* mains, and does not consider communication pipes. This is primarily due to the lack of sufficient data on these failures.

Many utilities are moving towards risk-based approaches for managing their pipe replacement and rehabilitation programmes and these approaches also include various methods for pipe deterioration modelling. Small diameter distribution mains are more likely to fail (due to the smaller wall thickness and structural strength). However, the consequence of a failure on a large diameter transmission main can be catastrophic. While sources of hydraulic unsteady conditions are often located at transmission mains (e.g. pump stations), there is plentiful evidence of significant pressure variations cycles observed inside the distribution networks. Besides, behaviour of particular assets inside distribution networks (e.g. large demand by an industrial consumer) can generate dynamic pressure variations that are travelling to even small diameter pipes towards end of distribution networks. Therefore, it is necessary to consider both small and large diameter mains.

With regard to pipe material, the primary target of this study is on the *metallic pipes*. Compared to non-metallic pipes, these form a larger proportion of the study networks. Metallic mains, specifically Cast Iron (CI) pipes, have been in service for a longer period of time and exposed to various deterioration mechanisms, therefore have higher density of pipe failures occurring on the metallic pipes.

A large proportion of the *data* used for this study are collected through the experimental programme designed and developed for this work. However, several datasets used in the analyses rely on the existing corporate data, including data on network assets and operations, hydraulic models, Geographic Information Systems (GIS), SCADA, etc. Several review and cleansing exercises are undertaken in order to validate the data, however gaps and inaccuracy in data could result in errors and should be carefully considered. The data checking process and any potential quality issues are discussed in the relevant sections where data is used, with an overview provided in chapter 3.

Many of the analyses undertaken in this work rely on information about the network's hydraulic operations. While *operational changes* across the supply network are common, it is assumed that no significant changes have been made to the study areas in a way that changes the system's hydraulic dramatically. Historical records on system characteristics and details of changes applied are often not available, making investigation in this area challenging. This assumption is confirmed by interviewing network's practitioners and operational management teams (Raine, 2014; Wilkie, 2014). This assumption is a research limitation for case studies where system hydraulics have changed significantly over time.

The *cost-benefit framework* intends to offer a general platform for evaluating cost effectiveness of calm network activities and innovations. Practical implementation of the framework relies on the data specific to the study network. A range of data and information is needed regarding pipe failure impacts and associated costs (direct, indirect, social, regulatory, etc.) as well as cost of the intervention activities and technologies for calm network. Such information is often case specific and driven by factors affected by the company's goals, vision, business plan, investment strategies, etc. Development of a detailed framework that considered localised data is beyond the scope of this study.

1.2. Novelty and contributions

Pipe failure is often the result of complex interactions between several factors. Much research has been undertaken in this area over the last many decades, developing the insight into this challenging phenomenon. This work aims to make contributions to enhance our understanding further and help providing a platform for future expansions.

The primary novelties and contributions from this study are to:

A. Develop a novel framework to investigate benefits of hydraulically calm

networks

B. Design and execute a large-scale surveying programme to gather network representative samples for development of data-driven statistical models

C. Formulate and develop a metric to quantify the stresses caused by hydraulic variations on pipes

D. Investigate causal analysis and burst prediction by developing statistical model to include the metric for hydraulic instability

E. Quantify impact of quasi-steady and unsteady pressure variations on pipe failure

F. Develop a methodology for investigating cost-effectiveness of innovation and intervention measures aiming to achieve calm networks

Further to these, there are additional novelties and contributions from this research work, are to:

a. Develop simplified models for hydraulic tracing and energy dissipation in unsteady pressure variation

b. Develop a burst data collection platform

c. Make recommendation on inclusion of pressure fluctuation stress metric in deterioration models

d. Make recommendation on instrumentations and monitoring level for investigating systems dynamic hydraulic conditions

Industry impact

A summary of the research impacts for the water utilities and the water industry in general are to:

1. Raise awareness of characteristics and impacts of network's dynamic hydraulic conditions amongst network practitioners, operational and senior management

2. Reduce pipe failures and subsequent impacts such as supply interruptions, discolouration incidents, customer complaints, etc. and increase network reliability

Table 1.1.: Research objectives, novelties and chapters where discussed

Objectives	Novelty	Chapter
Unique insight into the system dynamic hydraulic behaviour	Framework to investigate benefits of hydraulically calm networks	8, 10
Develop experimental programme of data collection	A large-scale survey sampling programme	6
	Data analysis to develop a metric to quantify pressure variations stresses	7
Investigate the impact of dynamic pressure variations	Statistical causal analysis and burst prediction model	4, 8
	Quantify impact of pressure fluctuations on pipe failure	7, 8
	Simplified models for hydraulic tracing and energy dissipation	7
	Burst data collection platform	6
Analyse pipe failure mechanisms and existing deterioration models	Recommendation on inclusion of pressure fluctuation stress metric in deterioration models	10
Cost-benefit framework	Cost-effectiveness framework for calm networks	10
	Instrumentations and monitoring level for investigating systems dynamic hydraulic conditions	10

3. Enhance networks operational performance, leading to improved regulatory compliance, Key Performance Indicators (KPI) and Outcome Delivery Incentives (ODI) (Ofwat, 2016)

4. Improve performance of pipe deterioration models

5. Contribute to investment decision processes and proactive pipe replacement programme

6. Provide framework for cost benefit evaluation of innovative solutions

Table 1.1 summarise the objectives of this work, novelties relevant to each objective and a reference to the chapter where each one is discussed.

Publication and research dissemination

Findings from the research work have been disseminated in several academic and industrial platforms. The publications from the work include:

- Rezaei, H. & Stoianov, I. (2015) An insight to dynamic hydraulic conditions in water supply networks. International Water Association Young Water Professionals. 2015 Glasgow.

- Rezaei, H., Ryan, B. & Stoianov, I. (2015) Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. Procedia Engineering. [Online] 119253-262.
  Available from: doi:10.1016/j.proeng.2015.08.883.

The works was presented in several industry events, including UK Water Industry Research workshop (April 2014, Glasgow), Institute of Water conference (July 2015, Cambridge), Bermad Pressure Transient workshop (October 2016, Reading) as well as within Imperial College London and Severn Trent Water to technical and management teams across the Company in numerous occasions.

1.3. Thesis structure

The structure of the thesis is illustrated in Figure 1.1. The thesis begins by providing an introduction to the research (this chapter). This states the research problem and the motivations driving the research work. It discusses the gaps in the knowledge and summarise the research objectives and questions the work is aiming to address and the novel contributions arising from this, considering research scope and limitations.

Chapter 2 presents a background review on pipe failure mechanisms, modes and factors contributing to pipe deterioration and failure. Deterioration modelling is discussed and an overview of various models in this area is presented. The chapter provides a review of pressure transient, its root causes and adverse impacts. The chapter concludes by providing an overview of the current state of the UK water industry and considers data availability and quality in this setting.

An overview of methodology, processes, techniques and variables used is presented in chapter 3. Details on data sources, quality and general availability and techniques to extract and manipulate data are also provided.

Chapter 4 first gives a review of Severn Trent Water network characteristics, and then presents the single-variate analyses of historical pipe failure records and factors contributing to the pipe failures.

Chapter 5 discusses fatigue failure mechanism, its characteristics and develops the cumulative pressure-induced stress index.

In chapter 6, design and development of the experimental programme are presented. It begins with techniques used to design the data collection programme and goes on to review the execution of the programme and presents the verification of collected data. The chapter conclude with details of the data collection for pipe failure events.

Review and analysis of high frequency pressure data collected in the experimental programme is presented in chapter 7. Methods to extract cyclic events from the pressure data and development of cumulative damage metric are discussed. After this, hydraulic tracing and energy dissipation between network's nodes are described, with techniques developed to estimate energy level across the network.

Chapter 8 describe the development of the statistical models for a case study and also Severn Trent Water networks, and implementation of cumulative pressure-induced stress index estimated for network nodes.

The cost-benefit framework for evaluating calm networks and cost-effectiveness of intervention innovations is described in chapter 10, with chapters 9 and 11 discussing the findings, finalising with conclusions, recommendations and discussions on the future works.

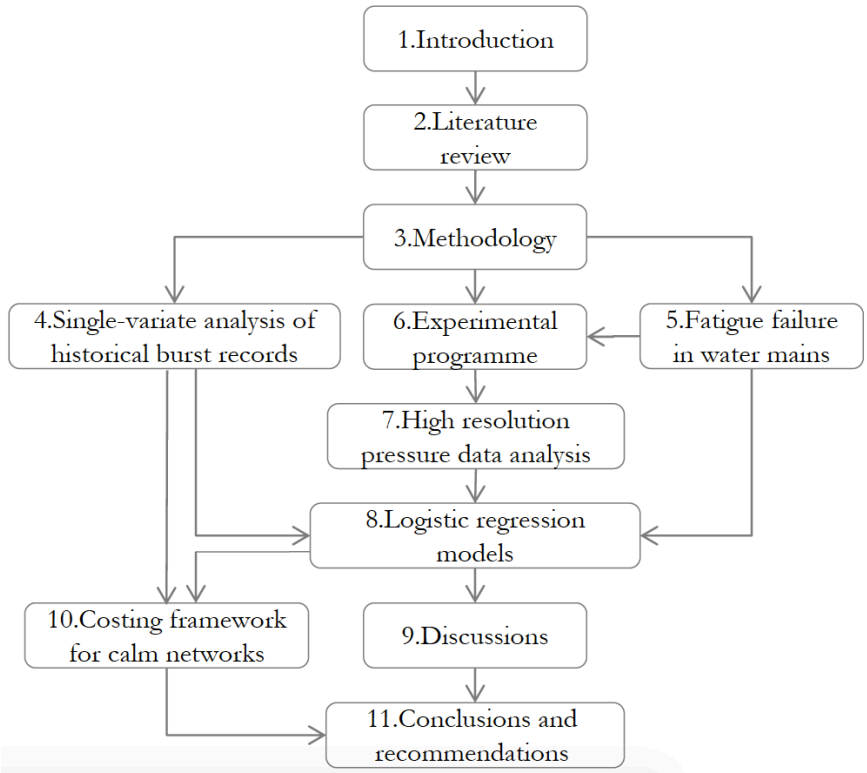


Figure 1.1.: Structure of the thesis with chapter numbers

1.3.1. Keywords

Throughout this thesis the terms *pipes* and *mains* are used interchangeably. *Pipe failure* and *mains burst* are both used and refer to a pipe repair event. Such repair activities might have been triggered by active leakage measures or as the result of a visible leak.

Furthermore, *trunk* main and *transmission* mains are used interchangeably. There does not seem to be a widespread agreements, in the academia and the industry, on the threshold diameter between transmission and distribution mains. In this work, the trunk/ transmission mains will refer to pipes with diameter larger than $250mm$.

The terms *hydraulic instabilities* and *pressure fluctuations* are used interchangeably and refer to quasi-steady and unsteady pressure variations.

Several acronyms and abbreviation are used in this thesis. The complete list of the abbreviations and symbols are presented in 'Abbreviations' and 'Symbols' sections on pages 7 and 10.

2. Literature review

2.1. Introduction

Failure of water supply pipes is the result of complex interactions between various factors related to the pipe, its environment and the loadings it is exposed to. Pipe failures have been studied from a fundamental structural deterioration perspective, analysing the impact of imposed loadings and pipe's structural behaviour, e.g. Kleiner and Rajani (2001), Makar et al. (2001) and Mohebbi et al. (2010). Studies have also examined relationship between pipe failures and various degrading factors, by employing statistical data driven techniques. Deterioration models are developed to determine probability and distribution of pipe failures.

Many studies in this area have focused on investigating the relationship between pipe failures and impacting factors related to the pipe and its surrounding environment. Pipe characteristics such as age and material as well as environmental factors, such as soil properties and temperature are considered in such analyses. Furthermore, stresses imposed from external factors, for instance soil movement and traffic loadings, have been analysed for their impact on pipes' deterioration.

Loadings from hydraulic instabilities and their impacts, however, have not been rigorously investigated. In recent years, few studies, such as Lambert (2001) and Lambert and Fantozzi (2010), have examined the relationship between system's maximum pressure and risk of pipe failure . These have also been limited to considering a single value (e.g. maximum pressure) as a hydraulic representation of the system. Therefore, most deterioration models developed based on these studies do not consider the dynamic hydraulic condition of the system and cyclic loadings as input factors to evaluate risk of pipe failure. The exclusion of this input variable is mainly due to the lack of high quality hydraulic data as well as tools and techniques to analyse and utilise high resolution data.

The aim of this chapter is to provide a review of the current knowledge in areas of pipe failure, deterioration modelling, adverse impacts of pressure transient events and the state of the UK water industry regarding pipe failure.

The chapter is structured as following. Firstly, a review is presented of pipes, pipe materials and various joints used in the water supply networks. There is then a review of mains failure, failure mechanisms, causing and/or contributing factors and resulting failure modes. Deterioration modelling, its concepts and existing models is discussed in the following section, followed by a review on pressure transient, its causes and adverse impacts. Finally, a brief overview of the state of the UK water industry regarding pipe failure and data availability and quality is presented. Gaps, limitations and shortcomings are discussed at the end of each section.

2.2. Mains and failure

Pipes and joints

The most popular pipe materials used in the UK water supply networks were Cast Iron (CI) and Ductile Iron (DI). *Cast iron* (also referred to as gray cast iron) is a strong but brittle material. Having been used at some places (e.g. palace of Versailles in France) for over three centuries, cast iron was first used in the UK (London) in the year 1746. Many cast iron water pipes used across the water utilities in the UK were produced some 50 to 150 years ago and manufactured in various ways from different compositions (Soil Pipe Institute, 2006). There are two types of cast iron pipe. Pit cast iron (gray iron cast vertically) and centrifugally spun cast gray iron, replacing the pit cast in the 1920's, with improvement in the tensile strength of the pipes, more consistencies in the wall thickness and use of cement for interior lining to prevent against corrosion.

Ductile iron pipe was developed from cast iron by changing the distribution of graphite in a spherical form instead of a flake form, resulting in vast improvements in fracture toughness and ductility as well as substantially more resistance to damage from impact or concentrated stresses (Makar et al., 2001). More resistant to corrosion and not subject to 'graphitic' corrosion, with greater strength and elongation compared to CI, DI was first used in 1948, with widespread use in 1960's. Such long history makes the metallic pipes the leading pipe material across many parts of the country.

Around half of the mains across STW are metallic and approximately three quarters of mains bursts are on metallic pipes (further details in chapter 4). While every network would have its own specific characteristics, due to its history, geography, population, land use, etc. STW network could be considered as a reasonably good representation of the UK water industry

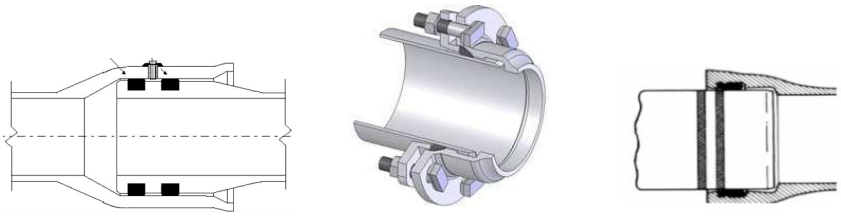


Figure 2.1.: Common joints used for CI and DI (a)Bell and spigot (b)Gland type (c)Tyton joint

networks. Fairly similar characteristics can be seen in water networks of other developed countries. According to Rajani and Kleiner (2001), in the US, more than two thirds of all water pipes are metallic with about half being cast iron and one fifth ductile iron. This is across all types of piping systems used by water utilities: transmission lines, in-plant piping systems, distribution mains and service pipes. This motivates the choice of focusing this study on the metallic pipes only.

Many of the pipe failures occur at the joints. Therefore, understanding joints and their characteristics is important, when investigating pipe failures. The most common joints used for both cast iron and ductile iron pipes are Bell and Spigot joints, mechanical Gland type joints, Tyton joints, mechanical lock type joints, mechanical push on type joints; some are illustrated in Figure 2.1. Bell and Spigot joint is where the spigot end of the pipe is inserted in the bell of the other pipe, with the space in between caulked with lead. The installation methods were very common until around 1935. After around 1920 until 1955, some utilities replaced lead with a sulphur compound, which resulted in more rigid joints contributing to cast iron main breakage. In 1950s, flexible rubber gasket joints were introduced (Lansey and Mays, 2000).

Mains failure

Due to variations in characteristics of buried pipes, properties of the surrounding environment and system operation and loadings, it is not surprising that pipeline failures exhibit a range of mechanical behaviour. Such behaviour is often the result of a complex interaction between various factors contributing to the mains degradation and ultimate failure.

Mains repair versus burst There are two stages in the development of a mains failure event. In the first stage, there is a burst or leak on the pipe but the pipe is still in service. If the leak is not identified, either through

active leakage management processes or when there is visible leak, this goes un-noticed and hence not treated as a failure. The second stage is when a leak is identified and/or the extent of leak becomes large enough to impact the network's hydraulics. This includes both burst opening scenarios, assuming a sudden pipe rupture or leak developing over a longer period of time. At this stage, activities are undertaken to repair the failed main. Depending on the water utility, a pipe burst might be defined as when a sudden rupture in the pipe results in significant change in the system hydraulics (e.g. supply interruption). In Severn Trent Water and most other UK water utilities, a pipe burst is defined when a mains repair activity is undertaken. This could be as the result of a sudden pipe failure, visible leaks or active/passive leakage management activities. This is the definition of mains burst used in this research work.

Assets life cycle

The life cycle of a typical buried pipe can be expressed by a 'bathtub' curve. There are two types of bathtub curves: one is the rate of occurrence of failure (ROCOF) of a repairable unit, and the second one is an instantaneous failure probability (hazard function) of a non-repairable unit (Ascher and Feingold, 1984). Water distribution networks pipes are generally repairable, and therefore the ROCOF type, shown in Figure 2.2, is considered most appropriate. There are three phases of a bathtub curve:

- The *burn-in* phase that is the early time after the installation, when breaks are not regular and happen mostly due to improper installation and/or manufacturing fault.
- In *normal life* phase, the failures are usually very few, occurring due to unusual factors, such as heavy loads and third-party damages.
- *wear-out* is the phase at which the failure rate is increased due to pipe degradation (Rajani and Tesfamariam, 2004).

These phases and their corresponding length may significantly vary from one pipe to another. As shown in the bathtub curve, due to pipe degradation, failure rate increases significantly during the wear-out phase. Such degradations are the result of the pipes' characteristics, age and other operational and environmental factors that impact the pipes' structural integrity. Due to their age, a large proportion of the metallic pipes in the UK are in the wear-out phase, reaching the end of their service life.

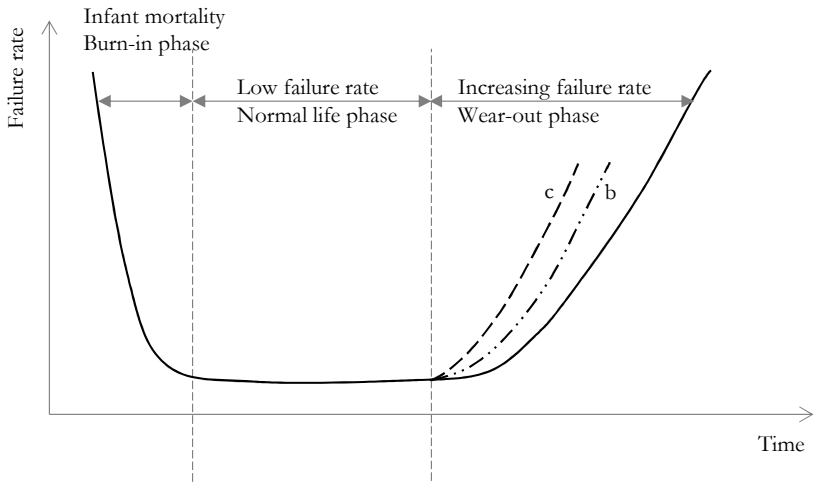


Figure 2.2.: Bathtub curve and shift in the wear-out phase due to additional stresses imposed on the pipe-adopted from Rajani and Tesfamariam (2004)

This research work formulates a hypothesis to investigate the impact of hydraulic instabilities in increasing the rate of failure in the wear-out phase, shifting the 'normal wear-out phase' and/or reducing the duration of this phase, thus accelerating the start of the 'burn-out phase' of a pipe, to steeper curves (curve b or c), as shown in Figure 2.2. Any structural degradation (e.g. crack initiation and propagation, stress corrosion cracking, etc.) as the result of cyclic loadings and dynamic hydraulic conditions in the system can lead to a reduction in the structural integrity of the pipes, until it reaches the point of breakage.

There are numerous factors contributing to pipe degradation resulting in a pipe failure through a variety of failure mechanisms, as discussed in the following sections.

2.3. Failure mechanisms and causes

At the most fundamental level, pipe failures are caused by applied forces exceeding the residual strength of the metal. Pipe breakage occurs when the operational and environmental stresses act on pipes where corrosion, degradation, inadequate installation or manufacturing defects has impacted the pipes' structural integrity (Rajani and Kleiner, 2011). The residual structural capacity of water mains is affected by material deterioration (due to environmental and operational conditions) as well as the quality of manufacturing and installation. The reduction in strength (e.g. as a result of corrosion) can be analysed on

the basis of a loss of pipe wall thickness (residual wall), or by using fracture mechanics, according to Belmonte et al. (2009). The initial strength of newly manufactured cast iron pipe is defined by either casting defects (e.g. porosity and inclusions) or by micro-structural defects. As the main ages, subsequent degradations result in the defects from corrosion governing the pipes' strength.

The physical mechanisms of failures in mains are usually a complex function of many uncertain contributing factors, related to:

- *Properties of pipe*: material, length, size, structural properties, manufacturing and installation quality, pipe-soil interaction, previous break and maintenance history.
- *Internal and external loadings*: due to operational pressure, as well as loading from surrounding ground, traffic loads, frost loads and third party impacts.
- *Material deterioration*: due to external and internal chemical, biochemical and electrochemical environment.

Table 2.1 lists some of the most important contributing factors to the pipe structural failures that are currently used in the state of the art deterioration models.

2.3.1. Loads on mains

Different types of forces applied to water pipes can be grouped as: bending forces, crushing forces, those produced by internal water pressure, soil movement-induced tensile forces and temperature-induced expansive forces. Sources of these forces are operational water pressure, ground loads (swelling soils and frost heave), pipe and water content weight, traffic loading and pressure transients.

Most loadings on pipes are non-static, often random in nature and could occur at various frequencies. The transient pressure wave event will typically reach peak pressure and attenuate over a few seconds. As for traffic loads, only the peak stress is treated as an event. Gray cast iron pipe systems were generally designed to withstand only internal pressure and crushing forces ground weight or truck loading above the pipe. However, pipes might be exposed to a range of cyclic loadings with varying amplitudes from the network's hydraulic instabilities.

Cyclic loading could lead to sub-critical crack growth from corrosion-induced, manufacturing and installation defects and also contribute to crack propagation

Table 2.1.: Pipe failure and the contributing factors

Contributing factor	Notes (from the literature and internal Company findings)
Pipe's diameter	Increase in failure rate as pipe diameter size reduces (Rajani and Kleiner, 2001)
Pipe's material	Material type influences corrosion and strength characteristics; Absence of external and internal lining and coating result in lack of protection from stresses; CI failure is generally circumferential, as a result of external corrosion; Spun CI has higher failure rate than pit CI
Pipe's age	Age is an underlying factor for the increased exposure to degrading environment and loading, methods of pipe manufacturing and installation; Ageing joints will increase burst rate
Pipe's length	Frequency of connections influences failure rates; Rate of increase in failure decreases with pipe length (curved relationship)
Ground conditions	Aggressive ground conditions and high shrink/swell factor accelerates degradation and failure in iron pipes; Backfill medium affects external corrosion rates; Contaminated soil may influence external corrosion; Depth of pipe burial and freeze/thaw susceptibility impacts pipes failure; Clay soils has greater frequency of bursts compared to chalk areas; Ground movement is particularly influenced by weather extremes and in clay soil leading to circumferential breaks; Lack of soil support where fine material being washed away can result in circumferential cracks in smaller mains and in pipe joints (de Oliveira, 2010)
Temperature/ weather conditions	Frost penetration results in ring load due to expansion of frozen moisture, mainly causing longitudinal breaks (Kleiner and Rajani, 2001); Burst frequency rises in late summer and culminate in the winter peak; Severe winters will exacerbate burst rates; Circumferential failures are greatest in winter; Smaller diameter pipes are more likely to fail in winter; Freezing conditions is not always the primary reason for failure (Stuart, 2011)
Corrosion (internal)	Internal corrosion is affected by the water pH, dissolved oxygen, alkalinity, temperature, chlorine, microbiological activity (Kleiner and Rajani, 2001); Changes in water treatment regimes may impact on the rate of internal corrosion
Corrosion (external)	External corrosion is a common cause of failure; External damage caused during pipe laying can influence speed of corrosion; Soil-pipe interaction results in deterioration and weakening of the pipe; Acids generated by degradation of organic material lead to corrosion; Soil with good electrical conductivity (e.g. clay) can cause corrosion in metal pipe; Corrosion is heavily affected by the lack of cathodic protection, particularly for transmission mains
Internal pressure	Longitudinal breaks occur specifically in larger mains as a result of excessive internal pressure loadings, specifically pressure transients
External loading	Loadings from earth surface (e.g. traffic) result in excessive stresses and ultimately failure of smaller mains; Heavy traffic conditions influence burst frequency

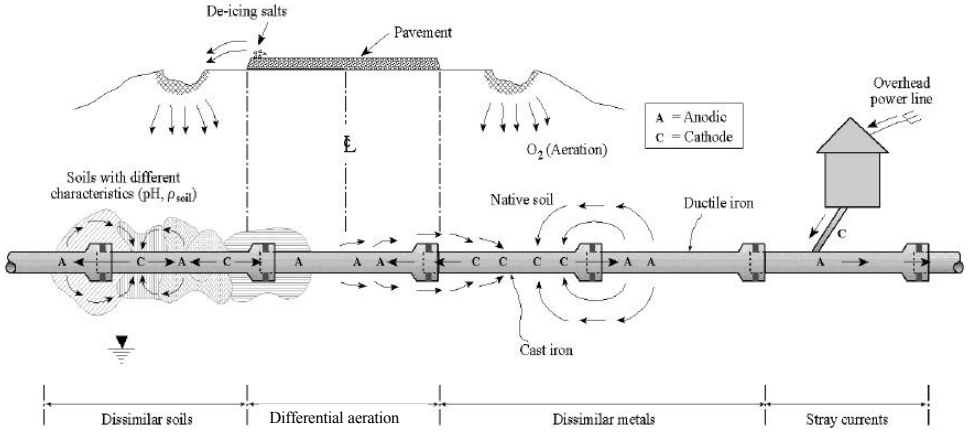


Figure 2.3.: Conditions leading to corrosion - adopted from Rajani and Kleiner (2001)

to eventually accelerate and result in a failure . In order to investigate and undertake a fatigue crack propagation analysis, it is necessary to determine the load history and consequent stresses. Measurement of the frequency at which each load occurs is used to generate the load history for random events. The load history could be complex and therefore may require to be reduced to a number of constant amplitude events, using a 'cycle counting' process. Details of the cycle counting methods are provided in Bannantine et al. (1990). For a failure to occur on a pipe, regardless of the course of loadings, the stress level must exceed the nominal material strength. On the other hand, as the service stresses are usually low, it is more likely that the development of a critical defect (e.g. as a result of corrosion) will lead to a degraded pipe strength (Atkinson et al., 2002).

Loadings from the network's hydraulic instabilities could adversely impact and exacerbate the effect of various underlying factors that causes pipe failure. Cyclic loadings from the network's hydraulic can contribute to the pipe's corrosion and crack propagation causing fatigue degradation and ultimate failure (see section 2.3.2 and chapter 5).

2.3.2. Corrosion

Corrosion is a physico-chemical interaction between a metal and its environment that leads to changes in the properties of metal. Corrosion is usually the main cause of pipe deterioration, common to all pipe diameter and failures due to this are very common (Makar et al., 2001). Figure 2.3 illustrates some of the conditions leading to pipe corrosion.

Pipe corrosion results in various problems. Firstly, as result of oxidation

and graphitisation, pipe mass is lost resulting in leakage and water quality issues. Furthermore, corrosion scales accumulate in irregular structures on the pipe wall which interfere with the water flow, generating additional headloss. Effective diameter of pipes could be reduced as result of advances in erosion, tuberculation and growing bacteria. Moreover, corrosion by-products (e.g. soluble iron) released into water impact water quality and aesthetic.

In grey cast iron pipes, there are two types of corrosion processes: corrosion pitting and graphitisation. Corrosion pitting is more common in larger diameter mains (above 300mm). In the graphitisation process, some of the iron in the pipe is removed, leaving a matrix of graphite flakes held together by iron oxide. In this situation, pipe may appear undamaged while there is actually no structural strength. Graphitisation is less common in DI, as there is no matrix of cast iron flakes in DI. Corrosion-induced failures originate either due to inherent defects in CI microstructure, intensified by corrosion or is because of pits developed in service as result of graphitisation (Atkinson et al., 2002).

Corrosion can occur both inside and outside of the pipe, resulting in weakening of pipe wall. External corrosion is the result of stray currents conducted between the pipe and the surrounding environment. Internal corrosion is mainly due to chemical reactions and interactions between the inner wall and substances in the water (Vipulanandan et al., 2011).

External corrosion

External corrosion has been identified as the leading factor contributing to metallic pipes' deterioration processes (Makar et al., 2001). There are four different forms of external material decay, namely: galvanic corrosion, acid attack, bacteriological action and electrolytic corrosion.

Galvanic corrosion is an electro-chemical process between two metals in electrical contact and immersed in an electrolyte, where one metal corrodes preferentially to another. Electro-chemical corrosion is shown to be the predominant cause of deterioration mechanism on the exterior CI and DI pipes (Rajani and Kleiner, 2011). In order for a galvanic corrosion cell to exist, four components need to be present; an anode (more reactive metal e.g. steel pipe); a cathode (more noble metal e.g. brass fitting); an electrolyte (a substance containing free ions, e.g. soil) and a conductive connection between anode and cathode. When two or more metal come to contact, there will be exchange of iron (more active metal to loose first and corroded). Cast iron, being made up of ferrous and iron, makes a perfect condition for galvanic corrosion. The surrounding soil could impact the cell development depending on its properties

such as moisture, level of aeration, acidity, pH, resistivity (i.e. soil resistance against flow of electricity), corrosivity, and difference in original and backfill soil surrounding the pipe. The soil maps released by the Soil Survey and Land Research Centre (SSLRC) details the extent of soil corrosivity across the UK (Cranfield University, 1997). Information from these maps is used to investigate the impact of soil corrosivity on historical main burst rate within STW areas, presented in chapter 4.

Acid attack refers to impact of soil's level of pH on pipes corrosion. Soils with pH values lower than 4 are acidic and can attack the pipe metal. These are also promoting galvanic corrosion by serving as a good electrolyte (American Water Works Association, 2003).

Bacteriological action is sulphide stress cracking experienced by most metallic pipes when in contact with solphidic environment such as moist hydrogen sulphide.

Electrolytic corrosion, also known as stray current corrosion, results from electrical current flow surrounding pipes, for instance, stray currents produced by tram lines, as concluded in the research published by Romer et al. (2005).

Internal corrosion

The interior of a metal pipe is impacted by various factors such as tuberculation, erosion, corrosion, and growing bacteria. Internal corrosion of pipes is caused by various factors including chemical properties of water, temperature, pH, dissolved oxygen, chlorine concentration, hardness of water (e.g. hard water-water containing calcium or magnesium salts is less corrosive), micro-organisms, flow velocity and operating pressure (Rajani and Kleiner, 2001).

Operating pressure does not necessarily initiate corrosion occurrence, however, it can have significant impact to amplify this. This includes the maxima of operating pressure, frequent cyclic pressure fluctuation and pressure transient events.

Stress corrosion cracking

Stress Corrosion Cracking (SCC) is the growth of crack formation in a corrosive environment; a phenomenon resulted from the combined action of both corrosion and, usually tensile stress.

There are three mains elements impacting the development of stress corrosion cracking; pipe material, corrosive environment and (above threshold) tensile stress. Therefore, when studying SCC, the following factors need to be considered (Ginzel and Kanters, 2002):

- Pipe coating and cathodic protection
- Soil condition, soil type, drainage, CO_2 concentration, conductivity, etc.
- Temperature (soil and pipe)
- Stresses: residual stress from manufacturing process; external stress: such as bending, welding and corrosion and secondary stress from soil movement
- Pipe pressure: crack growth rate increases with increase of (mainly hoop) stresses from the water pressure
- Cyclic loading effects (e.g. pressure, temperature).

When stress is applied to cast iron, the stressed area becomes more anodic, while the stress-free region tends to be cathodic, which accelerates the corrosion rate.

The type of stress corrosion cracking depends on the conditions that exist during the crack growth. For instance, as Ginzel and Kanters (2002) suggest, intergranular cracks are caused in existence of high pH, while non-classical SCC resulting in transgranular cracks are result of low pH.

Depending on the causes of crack, crack characteristics and profile can vary significantly. This might result in colony or isolated SCC. A combination of cracking and corroded area lead to a significant loss of wall thickness, as the depth of crack and corrosion are combined.

2.3.3. Joint failure

Joint failures usually develop in a gradual process and such failures are not catastrophic. Ground movement, pipe misalignment, movement of adjacent soil, poor pipe bedding and vegetation roots are some of various root causes for joint failures. In addition to these, cyclic loadings from traffic loads and pressure fluctuations can result in joint failure.

2.3.4. Ground movement

Often a combination of factors such as corrosion, the pipe's residual flaws and damages, and imposed stresses are causes of pipe failure. In absence of such combinations, a pipe failure might be associated with excessive forces applied to the pipe. A source of such forces is often ground movement, their origin are usually much further away from the failure location. Ground movements, as the consequence of swelling and shrinkage of soil due to its moisture content, are generally not uniform, because of variation in soil profile. Clay particles, the smallest particles of soil components, can affect the soil behaviour and movement, specifically when forming in excess of 35% of the soil components

and exposed to moisture. Due to its small particle size, the surface area of clay particles are very large and can attract a large volume of water. While clays' moisture content does not change greatly in their natural undisturbed conditions, causing large volume changes, when disturbed and exposed, the soil volume can increase and decrease significantly, depending on its moisture content.

A commonly used variable indicating soil movement is Fracture Potential. The soils in England and Wales are classified by the National Soil Research Institute (NSRI) based on their characteristics in shrink and swell (or fracture potential) from *Very Low* to *Very High*, as shown in Table B.1. Higher movement in soil will create higher pipe fracture potential, i.e. risk of failure.

Although there is a general agreement amongst practitioners that fracture potential is an important factor affecting pipe failures, previous studies have demonstrated conflicting results on relationship between fracture potential and pipe failure. For instance, Boxall et al. (2007) found no clear correlation between fracture potential and risk of failure.

2.3.5. Temperature effect

Temperature could impact the pipe breakage in two ways (Rajani and Kleiner, 2001). Firstly, this impact can be through temperature effects on pipe-soil interaction, and secondly, through frost load effect. Furthermore, temperature can play a major role in cast iron corrosion, through its influence on parameters that influence the corrosion. These parameters are fluid viscosity, dissolved oxygen concentration, physico-chemical properties of water, ion mobility, thermodynamic properties and biological activities. Corrosion rate is increased by three or four times as water temperature increases from 15°C to 60°C. On the other hand, in temperature below 5°C, cast iron and ductile iron materials develop embrittlement and therefore higher risk of brittle failure. Temperature changes in the environment surrounding the main leads to axial contraction and expansion of the pipe, resulting in added stresses to the pipe. The stress level would vary depending on the mains material. Cast iron has a different thermal coefficient of expansion, because of the multi metallic compound of this.

Frost heave As the frost continues to expand to the deeper ground layers, additional volumes of water freeze, resulting in extra vertical loadings to the buried pipes. The expansion could create significant stresses capable of moving and destroying buried pipes. Three factors are required in order for frost heave

to occur. Sub-freezing temperatures penetrating soil, a sufficient amount of water in the soil, and a frost susceptible soil capable of promoting capillary flow. Soil type, pipe depth, material and construction, as well as flexibility of above ground installation will define the susceptibility of a pipeline system to damage by frost heave.

Analysis of mains burst data undertaken by Rajani and Kleiner (2011) concludes that for cast iron pipes, there is an increase in burst numbers during the winter period. Further analyses on association of temperature and pipe failure is presented in chapter 4.

2.3.6. Pipe age and size

Pipe age has been shown to have a complex relationship with its probability of failure (Boxall et al., 2007). Pipe age is a surrogate for other pipe characteristics such as length of service, its material and strength as well as the manufacturing and installation technique used when the pipe was laid. Each of these impact the pipe performance and risk of failure. For instance, the majority of pipes laid before 1930's would be pit cast iron; pipes laid in 1920' and 1930's would be spun iron, significantly different in pipe's structural strength. Also, a pipe's length of service is an indicator of cumulative stresses imposed on the pipe as result of internal and external cyclic loadings as well as exposure to corrosive environment.

For pipe diameter size, in general, larger diameter pipes have a larger wall thickness and a higher moment of inertia. Therefore large diameter pipelines are less prone to failure compared to the smaller diameter ones. Smaller pipes are more susceptible to circumferential breaks, due to a lower moment of inertia, whereas larger pipes have a higher tendency of longitudinal break and shearing at the bell. Also, smaller pipes are particularly vulnerable to ground movement.

In their study, Andreou et al. (1987) concluded that smaller pipes break more frequently, experience primarily circumferential breaks (ring cracks) and are clearly affected by weather induced temperature changes in the soil (frost penetration). The predominant failure mode of larger diameters pipes is the occurrence of longitudinal breaks and no seasonality patterns (associated with weather conditions) are usually observed.

2.3.7. Other factors

Manufacturing flaws play an important role in failure of metallic pipes. Gray cast iron was manufactured in pit and spin casting. Porosity is the most common

manufacturing defect in pit cast pipe. In addition, unintentional objects created in the pipe during manufacturing can contribute to crack formation. Moreover, pipe wall thickness changes can weaken cast iron pipes. Often, pit cast pipes have non-uniform wall thickness. Presence of iron phosphide networks is another type of manufacturing flaws. Iron phosphide compound is more brittle than gray cast iron and weakens the pipe. Investigating the impact of *traffic loading*, Cooper et al. (2000), in a study conducted for Thames Water (UK), concluded that the probability of bursts increased with higher number of bus traffic. *Poor installation*, scratches to pipe coating during installation, poor bedding and backfilling and *design problems* are common type of human error resulting in increase in pipe failure rate.

2.4. Failure modes

Failure mode refers to the manner in which the pipes fail. Table 2.2, developed based on Vipulanandan et al. (2011), lists various failure classifications, failure modes and causes of failure.

Types of failures and their distribution would differ depending on characteristics of the distribution systems. In their study, Rajani et al. (1996) found that 70% of breaks are circular and 30% are longitudinal, holes caused by corrosion, or hydrant and leaks caused by faults in service connections.

Figure 2.4 illustrates the most commonly occurring mode of failure for buried pipes. In Figure 2.4-a, for small pipes (<380mm), on the left pipe a bell splitting is shown at the top and a circumferential cracking at middle of the pipe; where on the right pipe is shown corrosion through hole at top, chain created corrosion pitting at middle, elongated corrosion pitting with blow-out hole at bottom of pipe. In Figure 2.4-b, for large pipes (>500mm), left pipe shows a longitudinal splitting with the right pipe showing bell shearing.

Circumferential breaks

Circumferential breaks are usually due to thermal contraction (low temperature of water and pipe surroundings), bending stress (soil differential movement), inadequate trench and bedding practices, third party interference and axial temperature-induced pipe-soil interaction causing additional stresses on pipes. In their analysis, Kleiner and Rajani (2001) found that circumferential cracking is responsible for around 80% of failures in smaller grey cast iron pipes (<24 inch or 600mm). It is also highlighted that around 90% of circumferential failures are associated with the presence of corrosion pit at the failure surface.

Table 2.2.: Failure modes and classifications-based on Vipulanandan et al. (2011)

Types of failure	Failure modes	Causes of failure
Joint failure	Joint burst	Soil movement, thermal variation, expansion/contraction
	Connection failure	Welding defects, fatigue weakening, thermal stress
	Brittle failure	Graphitization, damage to coating, hydrogen embrittlement
	Bell split	Thermal expansion and contraction
Stress failure	Longitudinal/ split pipe	Ambient temperature differences, damages from transient
	Circumferential/ transverse break	Thermal/ circumferential/ mechanical stress, transient condition, soil movement
	Spiral cracking	Combined effect of bending forces
Corrosion and environment	Graphitization	Corrosive soil, anaerobic bacteria, stay current, hydrogen embrittlement
	Blow-out/ Pitting holes	Corrosive soil, stray current, microbiological influence
	Secondary effects	Chlorides from water, damage to coating, soil movement and dissimilarities, hydrogen embrittlement
Fatigue-induced	Corrosion fatigue	Combined effect of a corrosive environment and action of cycling stresses
	Creep fatigue	Combination of fatigue loading, when creep damage is accumulating
	Thermal fatigue	Prolong and repeated exposure of metal to cycles of heating and cooling



(a) (b)

Figure 2.4.: Failure modes for (a) small and (b) large diameter mains-adopted from Makar et al. (2001)

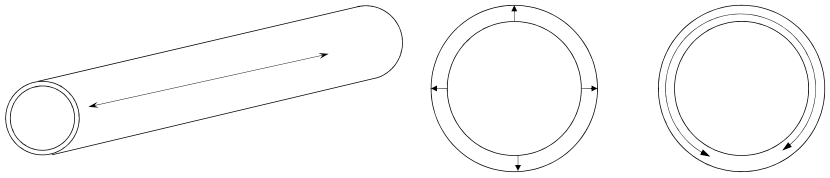


Figure 2.5.: Stresses imposed on pipe; longitudinal stress (left), radian stress (middle), hoop stress (right)

Longitudinal breaks

According to Rajani and Kleiner (2011), longitudinal breaks (caused by the transverse stresses) are typically the result of one or more of the factors below (as shown in Figure 2.5):

- Hoop stress due to pressure in the pipe
- Ring stress from the loading of the soil cover
- Ring stress because of loads caused by traffic
- Ring loads increase when expansion of frozen moisture in the ground is caused by penetrating frost.

Research by Makar et al. (2001) suggests that longitudinal cracking are relatively confined to large diameter pipes. Large diameter pipes generally have a high moment of inertia and therefore are less prone to circumferential failures.

Split bell

In bell split type of failure, a crack starts at the bell of a pipe and moves along its length before turning towards the pipe circumference. It is generally caused by thermal expansion or contraction. Results of a study by Makar et al. (2001) suggests that large diameter gray cast iron pipes can fail by having a section of the bell shear off. This could be due to compressive forces pushing the spigot of the pipe to the bell of adjacent pipe. However, bending forces are more likely to be the cause of this type of failure.

Spiral cracking

A combination of bending forces is the root cause of spiral cracking. In general, medium diameter pipes suffer from this. This is where the crack appears to start in a circumferential fashion and then develop along the length of pipe in a spiral shape.

Blow-out holes

In general, corrosion pitting only triggers the problem of blow-out holes but does not directly lead to pipe failure. However, Makar et al. (2001) argues that, in some cases, even corrosion pitting alone can still cause pipe breakage. This is where corrosion pitting develops until the pipe wall thickness is so low that water pressure blows out the remaining thickness.

Fatigue induced failure

Fatigue of cast iron has been studied extensively in structural elements, cast iron bridges and other disciplines such as automotive and aero-space. However, as highlighted by Rajani and Kleiner (2011), cast iron fatigue has not been adequately investigated in the context of buried water pipes . Failures due to fatigue occur when a pipe is exposed to repeated loads, which never reach to a high enough level to cause failure in a single occurrence. Fatigue cracking is one of the primary failure mechanisms in pipes. This is triggered by various factors such as corrosion, temperature, etc.

Fatigue processes in metal pipes can be divided into two phase: the *initiation* phase and the *propagation* phase. In the initiation phase, a crack initiates and grows into a small crack. This crack will then grow bigger, eventually leading to failure at the end of the propagation phase. The duration of these phases is expressed typically in stress/strain number of cycles, depending on variables such material, component geometry, etc. In addition to crack initiating in the pipes, manufacturing techniques employed during certain period of time ($\sim$1850's - 1930's) have been thought to cause voids and iron matrix in pipe structure, exposing pipes to development of fatigue failure, according to Rajani and Kleiner (2011).

There are different types of fatigue failures including *fatigue corrosion*, *creep fatigue* and *thermal fatigue*.

Corrosion fatigue is a result of the combined effect of a corrosive environment and the action of cycling stresses, where the total effect is greater than the effect from either of these alone. According to Leonard (2013), great variation and fluctuation of pressure loading, high enough peak stress to initiate crack, and sufficient cycles of stress are required in order for fatigue cracks to initiate. Environmental related factors include temperature, dissolved concentration, pH, chloride concentration, etc.

Fatigue failures originate from internal defects, such as corrosion, blow-holes, grooves, etc. Such defects could be caused by various factors. Once, due

to such defects, a pipe's strength is degraded, loadings imposed on the pipe can increase the risk of failure, as described by Belmonte et al. (2009) and Mohebbi et al. (2010). Areas with higher exposure to cyclic loadings are more prone to corrosion fatigue failures. This includes, for instance, regions of the pipeline system exposed to varying pressure and also urban pipelines because of major traffic loadings. Once a crack is initiated, it cannot be prevented from propagating. The process of corrosion fatigue is summarised by Baboian (2005) as: cyclic plastic deformation, micro-crack initiation, small crack growth to link-up, and macro-crack propagation. In corrosion fatigue (similar to pitting corrosion), the material as a whole remains unaffected and the region of attack is confined to a few points. Fatigue induced corrosion cause a number of cracks rather than impacting on only one point triggering the failure. Usually, surface defects, such as pits, are sites of initiation of cracks. Fatigue corrosion appears to cause less branching.

Fatigue failure on cast iron main can occur only if the pipe has existing cracks or voids (Rajani and Kleiner, 2011). This could be from manufacturing flaws. However, cracks can still be introduced during loading, installation, joining, etc. A pipe can have a long service life provided crack sizes are below a certain limit, and the pipe is exposed to low stresses.

Creep fatigue happens as a result of a combination of fatigue loading, when creep damage is accumulating. Creep occurs when material is exposed to high level of strains for a long time. For creep to happen, a combination of high stress and high temperature is ideal. Creep fatigue damage is more common at highly restrained joints.

Thermal fatigue is caused by prolonged and repeated exposure of metal to cycles of heating and cooling. It is rare in cast iron.

2.4.1. Gaps and limitations

Pipe failure is a complex phenomenon, often as a result of the combined impact of multiple, sometimes related, factors. Despite a large body of published material on the mechanism and causes of pipe failures, there are often difficulties in root cause analysis and investigation of the failure contributing factors. Due to limited access to buried pipes, gathering high quality data provides an additional challenge in investigating failure mechanisms.

Further complexity of the challenge arises from the ageing infrastructure being exposed to whole new ranges of hydraulic and operational changes, due to increased demands and also efficiency driven measures. Such operational

alterations will lead to additional cyclic stresses imposed on the network's assets.

Potential impacts of cyclic loadings on mains have been known for some time. As suggested by CIPRA (1927), some cast iron mains may have failed due to fluctuating and alternating stresses, for instance from surface traffic and unsteady operating water pressure and pressure transients events. Studies in this area have mainly investigated large diameter pipes exposed to significant cyclic loading (e.g. large pressure transient events). There is a clear need of a comprehensive systematic study of this phenomenon on various water pipes.

In order to develop better understanding of pipe failure at the network level, deterioration models are developed, as detailed in the following section.

2.5. Pipe deterioration modelling

Deterioration of pipes is categorised as structural deterioration and pipe's inner surface deterioration (Kleiner and Rajani, 2001). In structural deterioration, the pipe's structural integrity is affected, reducing its ability to withstand external stresses. Inner surface deterioration results in diminishing hydraulic capacity and reduced water quality. In this study, deterioration modelling refers mainly to the structural degradation of pipes ultimately resulting in pipe burst.

Deterioration modelling is often used to determine the mean number of potential failures, risk of pipe failure and distribution of these failures for specified time and space (Goulter et al., 1993).

There are two stages in pipe deterioration (Andreou et al., 1987). In the early stages of deterioration, pipes experience very few breaks with long time interval between failures. The primary importance in this stage is not the break-rate but rather the probability of a break in any given time period. In the late stages of deterioration, pipes will experience frequent breaks with no particular trend of increasing or decreasing break-rate with time, demonstrated by Andreou et al. (1987).

Undertaking deterioration modelling provides assessment of the rehabilitation, replacement programme and budgetary needs, cost effective inspections, problematic materials and operational areas, and associated depression (cost) of assets. Replacement programmes in the majority of water companies are based on previous bursts in the areas and complaints received from customers. Using this approach, many of the burst impacting factors will not be included development of such programmes.

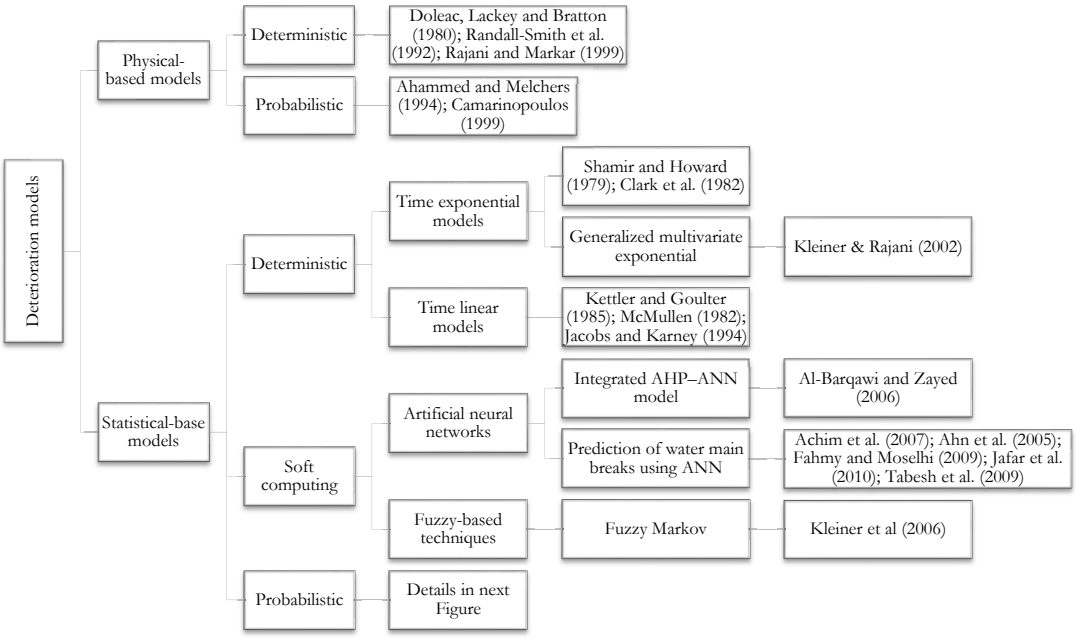


Figure 2.6.: Deterioration model categories and example models(1)

2.5.1. Deterioration models

In one categorisation, there are two approaches used to estimate pipe breakage vulnerabilities in water pipelines, as summarised by Kleiner and Rajani (2001): physical-based and statistical-based modes. *Physical-based* models are used for the prediction of remaining service lives of mains and *statistical-based* models are employed for interpolation of the whole area. These approaches can be applied in succession. While physically based methods are based on theories of mains failure (due to mains deteriorations) combined using mathematical relationships, statistically based methods only use mathematical equations with no account for natural failure mechanisms. Illustrated in Figures 2.6 and 2.7, there are various subcategories for each modelling approach and various models have been developed under each category.

Physical-based models

By analysing the loads the pipe is exposed to and the pipe capacity to resist these loads, these models attempt to predict pipe failure (Rajani and Kleiner, 2011). This approach requires understanding of degradation and corrosion processes in pipelines. They can be categorised into deterministic and probabilistic models.

In *physical deterministic models*, input variables have specified values and these values do not change temporarily or spatially. Such values are determined

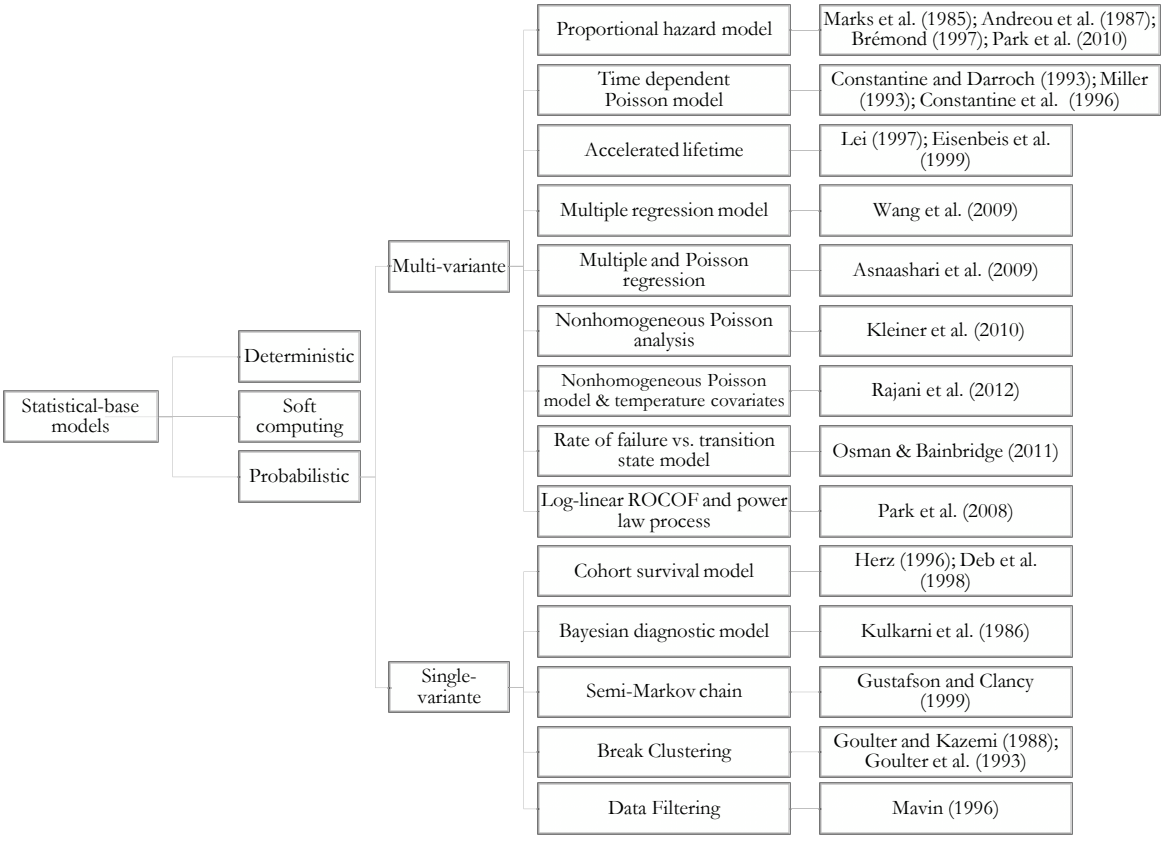


Figure 2.7.: Deterioration model categories and example models(2)

by lab tests or design standards. Common models used in this category are:

- (*Doleac, Lackey, and Bratton ,1980*) (Doleac et al., 1980) based on calculated remaining wall thickness, provide estimation for remaining service life.
- (*Randall-Smith et al.1992*) (Randall-Smith et al., 1992) based on pipe age, provide a linear model to calculate remaining life.
- (*Rajani and Markar,1999*) (Rajani and Makar, 2000) by considering changes in structural resistance capacity and variations in corrosion rate, this proposed a technique to estimate residual service life of cast iron pipes.

Physical probabilistic models consider various temporal and spatial variability causing a level of uncertainty during estimation. Also, non-destructive inspection techniques come usually with uncertainty in detecting corrosion pits. Various models based on this method are published in Ahammed and Melchers (1995), Camarinopoulos et al. (1999) and Kleiner and Rajani (2001).

Statistical-based models

Physically based methods require significant data and usually costly investigations on pipe's deterioration process. To avoid these limitations, statistical-based models are developed to predict break patterns and the level of deterioration from historical data.

These models are categorised to descriptive and predictive models (Andreou et al., 1987). *Descriptive* models are used for describing the number of breaks in each cluster of pipes and providing insights about failure patterns; without giving any information about individual pipe's failure behaviour. *Predictive* models are used for predicting the probability of pipe failure by collecting past pipe breakage behaviour. Predictive modes are categorised as *aggregate type models*, *regression type models*, and *probabilistic* or *choice models*.

In *statistical deterministic* models, cumulative pipe breaks is expressed as a function of pipe age by exmploying various time-dependent equations. In their review, Kleiner and Rajani (2001) suggest two types of deterministic models: time exponential and time-linear models. Regression analyses used by Shamir and Howard (1979) to predict pipe's breakage is a time exponential model. Similarly, numerous studies have been carried out and various models have been developed for time linear models. Examples of such models have been presented in works published by McMullen (1982), Kettler and Goulter (1985) and Jacobs and Karney (1994).

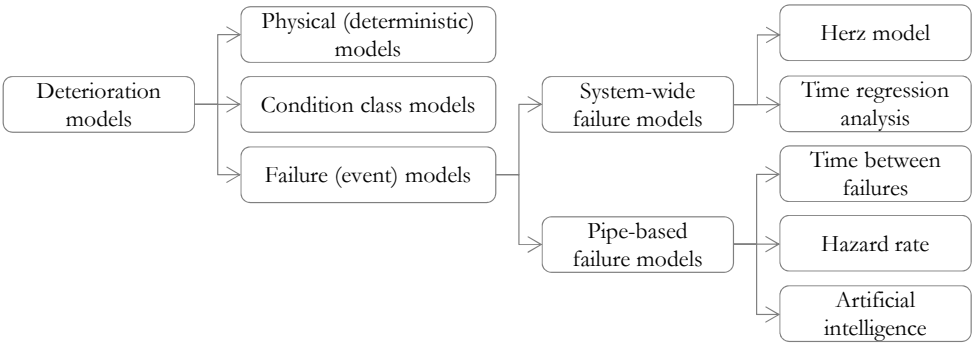


Figure 2.8.: Other categorisation of deterioration models

Statistical probabilistic models can be divided to single-variate and multi-variate models. *multi-variate* models estimate the failure potential as well as the distribution of failure. These models are useful when various covariates are considered in the analysis. This provides estimation of probabilities of bursts for particular pipes and also the difference between variables to be compared. Two types of regression models have been identified by Jianhua and Sveinung (1998): proportional hazards models and accelerated lifetime models. Probabilistic *single-variate* models determine burst probabilities on grouped or clustered data. The probability of pipe failure is related to the probability of observing specified characteristics on the failed pipe; the probability of observing the same characteristics on a pipe not failed, and the probability of failure.

Statistical Soft Computing or data mining models are seen as alternative methods to precise computing used in other statistical models. This approach is tolerant to uncertainty and imprecisions and can exploit the tolerance of partial truth and manage uncertain and imprecise information (Zadeh, 1994; Kleiner et al., 2006). Fuzzy techniques, artificial neural networks, genetic algorithms, probabilistic and evidential reasoning are amongst the techniques in soft computing approach (Nishiyama and Filion, 2013).

Other categorisations

In addition to above, another way to categorise existing deterioration models is to group the models as physical models, condition class models and failure models (Figure 2.8, as suggested by Maurer (2013)).

Condition class model are very sensitive to missing historic data. However, these are very popular because of medium data demand and easy analysis of the available data. Examples of such models would be *Herz function* as well as other popular survival functions such as *Exponential, Weibull, Gompertz.*

There are two categories in *failure (event) models*: system wide failure rate models and pipe based failure models. *System wide failure rate models* are not useful for estimating the location of the next break, but can predict the number of breaks per unit length per year. An example of this is *time regression analysis* which is an exponential regression model. *Pipe based failure models* are very robust with respect to lack of data and popular for water mains. They can be used for prediction of the probability of a water main breaking and also for detailed rehabilitation planning. Various methods in this category are *time between failures*, *model hazard rate*, and *artificial intelligence models*.

Data requirements

Some of the main challenges when developing and using deterioration models (specifically for statistical models) are related to the available input data. Most models rely on quantity and quality of input data, that of assets, environmental factors and historical pipe failure records. Factors contributing to the shortage of quality data include age of infrastructure (missing data for assets installed long time ago), inaccessibility to buried pipes, inaccurate or missing asset data and undocumented alteration to assets. Furthermore, often, there are various techniques developed and used to quality control the available data and populate missing information. Further discussion on data availability and quality in the UK water industry is presented in section 2.7.1.

Different deterioration methods consider variety of factors that impacts pipe deterioration. Table 2.3 summarises the data requirements for some of the most well-known models.

Deterioration modelling in STW

Figure 2.9 illustrates a conceptual overview of deterioration modelling practice as currently undertaken in STW. The deterioration model is driven by the expected *outcomes* (including main renewal, lower leakage, and supply demand balance). There are various *constraints* which shape the working behaviour of the model, e.g. available headroom, supply and demand, level of investment, burst and supply interruption targets, ancillary repair needs, discolouration complaints targets, etc. are amongst these. Burst rate and supply interruptions targets are 'to be kept flat, i.e. cannot rise'.

Asset Data (including mains length, material, diameter, age as well as surrounding soil, and location) will be used to model pipe burst. These will define the Main Renewal planning. This will have a two-way impact on *interruption* data (material, location, diameter), and also *leakage*. Elements

Table 2.3.: Data requirements for various models

Model/ authors	Physical properties							Environmental factors							Mechanical properties						
	Pipe age	wall thickness	Pipe diameter	Pit depth (external)	Pit depth (internal)	Breakage history	Redox potential	Soil pH	Soil aeration constant	Soil resistivity	Soil moisture	External loads	Corrosion rate coefficient	Air temperature	Operating pressure	Velocity of fluid	Water temperature	Stresses	Tensile strength	Toughness	Rupture modulus
Physical deterministic methods																					
Doleac (1979) and Doleac, Leakey, and Bratoon	✓	✓	✓					✓	✓	✓											
Rajani and Makar (1999)	✓	✓	✓	✓	✓			✓	✓	✓		✓							✓		✓
Randall-Mith et al. (1992)	✓	✓		✓	✓																
Physical probabilistic methods																					
Mohammed and Melchers (1994)	✓	✓	✓						✓		✓	✓		✓	✓	✓	✓				
Camarinopoulos (1999)	✓		✓									✓	✓	✓	✓		✓	✓	✓	✓	
Statistical deterministic methods																					
Shamir and Howard (1979)	✓					✓															
Kettler and Goulter (1985)	✓		✓			✓															
McMullen (1982)							✓	✓		✓											
Jacobs and Karney (1994)	✓					✓															
Probabilistic multivariate models																					
Andreous and Marks (1985)	✓		✓			✓		✓	✓	✓		✓			✓						
Lei (1997)	✓		✓																		
Probabilistic single-variate models																					
Herz (1998)	✓					✓															
Kulkarni (1986)		✓				✓		✓	✓	✓	✓										

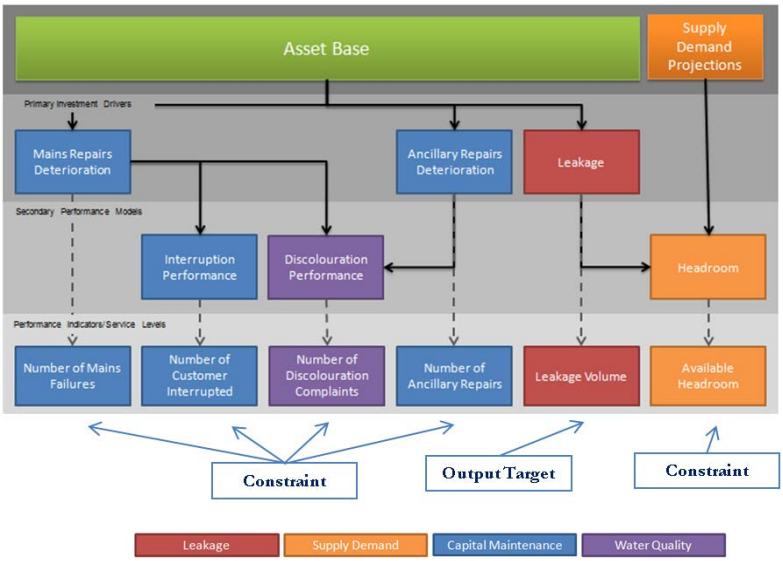


Figure 2.9.: Overview of the deterioration model at STW

related to leakage control and impacts include Active Leakage Control, Pressure Management and Main Renewal. This is in addition to Natural Rate of Rise. There are various measures to control/impact the Supply Demand Balance, including demand leakage, new resources, water efficiency measures, and metering. These elements have interdependencies with other elements of the model, e.g. leakage level, etc.

STW's deterioration model has been developed and continuously upgraded with new available data as well as new performance requirements from within the Company and external regulators. The model, however, is currently delivered in a commercial product with a '*black box*' approach, and hence it is not possible to investigate how the burst prediction model works.

Shortcomings of deterioration models

Pipe failure is a complex phenomenon, often the result of several factors' interaction and interdependent impacts. Deterioration models developed to study these often fail to explain pipe failure causality and/or to predict risk of pipe failures.

There are several challenges in development of deterioration models. These include technical challenges to choose appropriate model and datasets, challenges in choosing factors to consider, handling issues with data availability and quality, and expertise to develop and implement such models, etc.

Deterioration models often fail to consider all these factors and/ or their combined impact. As per Table 2.3, only a few current models consider

hydraulic condition (pressure and/or velocity) of the system as an impacting factor; and all ignored impact of (steady and unsteady) pressure variation. Furthermore, limited deterioration models that consider maximum pressure as a contributing factor rely on low resolution (15-minute interval) pressure data collected across the network. Such logging set-up can fail in capturing the maximum pressure and dynamic conditions of the system, specifically when pressure changes are rapid.

With advancement in data collection, handling and analysis technologies, new deterioration models can be developed to benefit from these development. Importantly, these model should consider other potential factors, such as pressure variation stresses, that can contribute to pipe degradation and failure.

2.6. Pressure transient

A hydraulic transient occurs following a sudden change in flow condition. This condition results in generation of pressure waves that transmit the energy from one point to another through the medium without significant displacement of the matter in between (Walski et al., 2004).

Flow in pipes is often assumed one-dimensional and incompressible. This, however, could be misleading when studying unsteady flows, specifically (rapid) pressure transients. Changes in the flow occur after a short interval. Such time interval depends on the distance from the point where change is initiated as well as the speed of pressure transient wave propagation. Therefore flow characteristics (pressure and velocity) at each point would depend on time (after a transient event occurs) and place (distance from the source of the event). As highlighted by Thorley (2004), speed of propagation depends on physical characteristics of the water and pipe.

A *rapid* pressure transient is developed when an event (such as pump trip, valve closure, etc. that originates the pressure transient event) is completed within a Pipeline Period, i.e. the time taken for the pressure wave to complete a return travel of the pipeline. This depends on the length of pipeline and speed of wave propagation. A method of protection against generation of a high energy pressure transient event is for valve or pump closure time to be greater than the pipeline period.

Change in the pressure as the result of pressure transient events can be expressed by the simplified but well known Joukowsky equation:

$$\Delta H = -\frac{c\Delta v}{g} \tag{2.1}$$

where c is the wave celerity or propagation speed (m/s), H is the head (m), v is the flow velocity (m/s) and g is gravity acceleration (m/s^2). c, can be calculated by:

$$c = \sqrt{\frac{1}{\rho[\frac{1}{K} + \frac{1}{A}\frac{\partial A}{\partial p}]}} \tag{2.2}$$

where ρ is density (kg/m^3), p is pressure (N/m^2), A is cross sectional area (m^2), K is bulk modulus (representing elasticity of fluid, N/m^2), and term $\frac{\partial A}{\partial p}$ representing elasticity of the pipe. Joukowski equation considers frictionless fluid and elastic pipe material. This approach results in several limitations in capturing the characteristics of the networks where friction and non-elastic behaviour of the system play important roles in defining the transient behaviour.

Equations describing a pressure transient phenomenon include the basic conservation of mass and momentum. In their general form, these are non-linear partial differential equations:

$$\frac{\partial Q}{\partial t} + gA\frac{\partial H}{\partial x} + R_f Q|Q| = 0 \qquad \textit{continuity} \tag{2.3}$$

$$\frac{\partial H}{\partial t} + \frac{c^2}{gA}\frac{\partial Q}{\partial x} = 0 \qquad \textit{momentum} \tag{2.4}$$

where A is the cross sectional area, g is gravity acceleration, c is the wave propagation speed, R_f is frictional resistance, H is piezometric head and Q is flow.

As a closed-form solution is not possible, numerical methods are used to solve these. *Method of characteristics* is the most popular technique, where the equations above are converted to ordinary differential equations and then solved by a finite difference method. Other methods of analyses of pressure transient include graphical method, algebraic method , implicit method, linear methods (Wylie et al., 1993) and Perturbation method (Walski et al., 2004).

The unacceptable conditions occurring due to pressure transients, Thorley (2004) argues, depend on the pressure changes amplitudes, speed of occurrence and length of time the system components are exposed to these. The amplitude of pressure transient events are impacted by: pipeline length and configuration, flow rate of change, elastic properties of water and pipe, contents of dissolved or gaseous gases, vapour pockets and protective measures in place.

Practical methods of pressure transient suppression are, as listed by Thorley (2004), direct actions and diversionary tactics. *Direct* actions are activities containing the increased pressure and changes in causes of pressure transient event. This includes employing stronger pipes, re-routing, changing valve

movement, avoiding check valve slam, increasing pumps' inertia and minimising resonance hazards. *Diversionary* tactics are more common and is based on releasing the surplus energy from the system. This method includes use of air vessels and air cushion chambers, accumulators, surge shafts, surge tanks, air release valves, pressure relief valves and bypass lines.

The energy of the pressure transient would gradually dissipate due to various dissipation mechanisms. Six primary mechanisms for dissipating transient energy are identified by Karney and Filion (2003) as:

- Convectional fluid friction
- Networks demand: increase in demand/flow in response to increase in pressure provides an effective dissipative outlet for transient energy.
- Leakage: this will provide a fairly similar energy outlet as the network demand, although unwanted and un-designed.
- Network characteristics: looped system, branch and service connections, junctions and isolations valves, etc. fragment pressure waves to smaller pieces, accelerating the energy dissipation.
- Unsteady friction: rapid acceleration of fluid causing distortion of velocity profile, hence high rates of energy dissipation.
- Specialised appurtenances designed in the system for the purpose of dissipation: such as air-vacuum valves, etc. This is different from all other mechanisms, as it is designed for the system, rather than an intrinsic part of the network.

Several methodologies exist in the literature to study the energy dissipation based on various dissipation mechanisms and considering network's characteristics and complexity. Further details are discussed in section 7.5.

Causes of pressure transient

There are various factors causing varying velocity that result in development of pressure transient events. Rapid opening and closing of valves, rhythmic valve operation, pump switches, movement of air pockets, sudden release of air, recombination after water-column separation, etc.

Sudden changes in operation of operational assets could be significant sources of dynamic hydraulic behaviour in the system, for instance as a result of pump/valve operations. Pumps switches, whether planned switch for operational purposed or unplanned switch for instance in an event of power failure,

could cause large scale instabilities in the system. This could be the same for valves, specifically larger ones. Changes in status of valves (either automatically controlled valves, such as flow control valves or pressure reducing valves, or manually operated valves) are significant sources of pressure transient events.

Furthermore, a sudden change in demand in the system, (for instance a large industrial consumer) could result in a significant change in system flow (and hence velocity) and therefore generate potentially very frequent and large transient events. In addition to these, other factors such as air valve slam, burst mains, etc. could be contributing significantly in generation of instabilities in the system.

One of the important, yet less explored, causes of pressure transient is operational activities in the system, undertaken by system operatives. When, either as part of normal maintenance or in an emergency, certain network interventions (such as opening/closing valves) are carried out on the network, if not undertaken correctly, these actions could cause alterations in system flow regimes, resulting in large instabilities and pressure transient events. There have been previous studies, for instance the findings presented in Creasey and Lynn (2011) from Water Research Centre and also internal studies at STW, looking into the impact of operational activities (mainly valves opening and closing manually undertaken by operatives) on creating bursts in the system. The STW's internal study, for instance, found that many of the bursts which occurred over the specific time period of the study had some operational activities in preceding days and as such concluded some of the bursts were potentially, to various extend, impacted by such activities. As a result of this internal study, a training rig was designed and developed by STW in order to train all system operatives on the correct procedure to operate valves, in an attempt to reduce potential impacts on the system.

Another significant factor contributing to this issue could be the drive towards efficiencies. While this, when implemented correctly, is a good and necessary approach, careful consideration will need to be paid to ensure the full implication of these measures is well investigated and understood. Many of the optimisation and management measures seek, often aggressive, changes in operation of the network assets in order to achieve efficiencies in resources (water, energy, etc.) consumption. Such changes, in most cases, mean more frequent operation switches of pumps and valves, which as discussed, could be the source of even more instabilities in the system.

Gaps and limitations

Pressure transient events and their impact on pipe failure are well studied. This is, however, limited to events of significant amplitude and usually impacts on individual large diameter mains and trunk mains. Adverse impacts of hydraulic instabilities at the network level and for smaller distribution mains are not given adequate attention.

A recent large-scale study, entitled *Critical Pipes*, was undertaken recently to investigate the pipe failure and factors contributing to the failure. This was a collaborative project between various water companies and universities in Australia, with (limited) collaboration from the utilities in the UK and north America. One of the main findings of the project was to highlight the importance of pressure transient in pipe degradation and ultimate failure, as reported by Rathnayaka et al. (2016) and Rajeev et al. (2013). These studies, however, did not provide a quantitative investigation of the significance of the impact.

A network's dynamic hydraulic conditions and deterioration of its assets are often quite complex to investigate. Furthermore, there are limited techniques to categorise ranges of dynamic conditions with varying characteristics and quantify the deterioration caused by these.

A significant gap in this area is lack of systematically collected high quality pressure data. Most pressure-focused investigations are currently undertaken using traditional low resolution pressure loggers (15-minute interval). However, these are not suitable for capturing rapid changes and instabilities in pressure. Traditionally, a lack of suitable recording, handling and analysing tools have been the main barriers in data collection. Recent development in instrumentation and analytical capabilities have made significant improvement in performance of data management platforms. The InfraSense platform developed at Imperial College London, is an example of this, providing a high quality data collection and processing tool and equipment.

Furthermore, systematic data collection is required in order to provide a network wide insight into the behaviour of hydraulic instabilities and the impact of network characteristics and assets on their propagation. In addition, detailed investigation is needed in analysing the data collected and quantifying characteristics of various events and their impact on network's assets.

There are several potential adverse impact from pressure transient that have not been investigated to a sufficient depth. For instance, the impact of pressure transient events on water quality (discolouration) incidents, particularly transient events that do not result in a pipe failure. It is possible that

sedimentations in pipes are disturbed and suspended in water, if shear stress resulted from pressure transient events results in dislocation of the sedimentation, causing discolouration events. An experiment was designed and undertaken in this study to investigate this. Several pipes, with potential risk of pressure transient events (e.g. downstream of large pump stations) were logged using high frequency pressure transient loggers. Simultaneously, these pipes were logged using turbidity monitors. The result from the experiment was not conclusive to prove or reject the hypothesis, mainly due to the following. Firstly, the resolution of turbidity monitors were to collect one sample at 5-minute intervals. In addition, the turbidity loggers available to this project were large bulky equipments, and therefore cannot be deployed on site (only used when operative is present). The data collected was not sufficient to investigate potential correlations with fast moving pressure transient events. Furthermore, as the pipes in the study had been exposed to repeating occurrence of pressure transient events, it is possible that any sedimentation that can be disturbed by shear stress from the regular pressure transient events have already been suspended. It is, therefore, necessary for a larger event to occur, if a new sediment dislocation is to happen. This remains an important area to be investigated further.

2.7. Current state of the UK water industry

In order to understand the current state of industry practices with regards to pipe failure and impact of pressure-related stresses on these, a series of meetings, discussions, collaboration and brain-storming activities were undertaken with practitioners across the water industry. The project's activities and goals were presented at a number of industry events, including at the UK Water Industry Research (UKWIR) and Water UK seminars.

In April 2015, a workshop was co-organised by the author and the Water Distribution Infrastructure Network (WDIN), a Water-UK group with representative from most water utilities in the UK, facilitating inter-company collaboration. The motivation for organising this was to validate a number of research hypotheses with other water utilities, including an industry wide lack of realisation regarding pressure transient impacts on networks' assets and specifically a systematic quantitative measure of adverse impacts from hydraulic instabilities. The workshop objectives focused on understanding the relevant practices across other major water companies in the UK. A set of questions were sent to the practitioners in advance of the workshop, followed by a presentation from each utility on the state of their current activities and

future plans and ambitions in this area. Following to the workshop, a protocol was developed by the author to develop a framework for high frequency pressure data collection and case studies at the participating utilities. This is intended to enable future research and developments in this area based on (larger and varied) data and information from several companies.

Findings from the workshop and other collaboration activities illustrate that there is a clear realisation being formed across the industry about the adverse impact of pressure transient. It is highlighted that most practices, however, tend to focus on short term solutions and a detailed systematic and rigorous investigation is required to study medium and long term impacts.

2.7.1. Data availability and quality

The level of availability and quality of datasets across the water industry often present significant shortcomings specifically when data-driven techniques and analyses are undertaken. Despite numerous initiatives to address the challenge, the issue is still particularly significant in data on infrastructures, buried assets and also historical data. With recent progress in data collection and storage facilities, the quantity of available data has improved greatly, however, there are still significant challenges to overcome in this respect. Outcomes of data-driven analyses should be interpreted with careful consideration regarding potential shortcomings (both quality and quantity) of the input data.

In undertaking this project, various tools and techniques were employed to ensure a careful review of the data. Data was gathered from a variety of available sources (e.g. GIS, corporate databases, etc.) followed by cross checks and validations to ensure data accuracy and reliability. Furthermore, spatial tools were employed to investigate associations between the existing data and information obtained from new sources (e.g. hydraulic models, field collections, etc.). For instance, corporate databases on association of each pipe segment to its DMA were cross checked by GIS-enabled spatial analyses and hydraulic model output. Further details on treatment of missing data is presented in section 3.3.1.

For the purpose of demonstration, two examples of data inaccuracy and techniques undertaken to address them are presented below.

Example 1 Longitude and latitude coordinates for a number of burst points are recorded incorrectly, as shown in Figure 2.10. Around 600 number of burst records ($\sim$%1 of total) have location information that are $50m$ or more distance from the closest pipe. Also, there are approximately 2600 (%3.7) bursts with identified locations of more than $5m$ from the nearest pipe. For



Figure 2.10.: Examples of data inaccuracy - distance of recorded burst location to the closest pipe and potential allocation to multiple pipes

Table 2.4.: STW pipe with missing data in the Year 2008

Attribute	Length(km)	Length(%)
Material	10931	21
Diameter	1918	4
Year Laid	36557	68

records with small errors, where possible, other information about the burst record (for instance material and size of the repaired pipe) is used to find the correct information. If reasonably confident allocation of bursts to a pipe is not possible, the records are removed from the data processed for analysis.

Example 2 The corporate databases on infrastructure and buried assets, (sometimes the only available source of data), have missing data for some key attributes and fields. Table 2.4 shows the lengths of pipes that did not contain data in certain fields in year 2008. After a detailed review of this, various techniques were undertaken by the Company in order to populate the missing data with the best estimate. For instance, using the information from the neighbouring pipes, the development period of the area where the pipe is laid, etc. missing information were populated.

While every attempt is made to ensure the available data is used with ultimate care, the underlying issues with quantity and quality of the data

should not be ignored. This poses the need for various assumptions as well as limitations on the quality of outcomes from any analysis based on such data, and therefore any use of such outcome should take into account such assumptions and limitations. In section 3.3 a review of the data used in this study, their sources and development as well as their general quality are provided.

2.8. Summary

In this chapter, a brief background review to the pipes' degradation, failure modes and mechanisms is provided. Pipe failure can develop as the result of complex interactions of various factors impacting pipe's structural integrity. Such complexity, in addition to various limitations in available high quality data and system hydraulic understanding, has lead to, sometime significant, shortcomings in understanding the root causes of pipe failures.

Deterioration models, and specifically statistical ones, are developed in order to provide a pipe and/or network wide estimation of probability of pipe failure, relying heavily on assets' and networks' data. There are shortcomings in the capabilities of these models, due to the lack of quality data and also exclusion of important factors, such as pressure instabilities and their loadings across the network. Pressure transient events, as result of rapid changes in network's hydraulics, can impose significant loadings on network's assets. An increasing body of evidence shows the importance of considering pressure fluctuation and hydraulic instabilities as key factors in pipe deterioration and ultimate failure. Few recent deterioration models have considered maximum pressure in their analysis as a factor impacting pipe failure. None has, however, considered network-wide impacts of pressure fluctuations, cyclic loadings and structural degradation imposed by these.

There is a clear realisation of importance of hydraulic instabilities in pipe deterioration studies, both within academia and practitioners, highlighting the need for a systematic investigation into behaviour of this phenomenon and its adverse impacts. The investigation requires design and execution of high quality data collection and processes, categorising and quantifying the instabilities events, followed by a detailed analysis on relationship between the hydraulic dynamic conditions and pipe failures. Undertaking prevention and remedial activities to mitigate impacts of hydraulic instabilities is often driven by analysing costs of such activities versus benefits gained from reduction in number and extend of pipe failures. Pressure transient intervention thresholds are required to assist decision making and investment planning.

In the following chapters, detailed methodology for investigating the research questions are provided, followed by single and multi-variate analyses of historical pipe failure at STW and design and development of an experimental programme to collect high frequency pressure data. The data is then analysed and statistical models developed to investigate relationships between hydraulic instabilities and pipe failure. Finally, a methodology is developed to quantify costs and benefits of intervention activities aiming to calm the network.

3. Methodology

This chapter provides an overview on how the research was carried out, data attributes used for different analyses, their sources and data manipulation processes. The chapter also describes the steps undertaken in collecting required data for the research and the techniques used to analyse the data and develop the statistical models. There are also references are made to various sections and chapters, where details of specific analysis or data collection are described.

3.1. Introduction

The overall aim of this research work is to investigate impacts of dynamic pressure variations on water pipes' structural failure. In this chapter, firstly, a brief overview of steps to the study is presented. After that, details are provided on how the research was undertaken, characteristics and sources of data used in each step, equipment used for data collection, concluding with discussion in analytical framework and statistical treatment of the data.

3.2. Overview of research development

A summary of the steps undertaken to investigate the research questions, detailed in chapter 1, is as following:

1. Reviewing pipe failure mechanisms and factors affecting pipe deteriorations
2. Undertaking single-variate analysis of historical pipe failure records and deteriorating factors
3. Designing and executing of sampling programme to collect high resolution pressure data
4. Analysing collected data and developing cumulative damage index
5. Developing multi-variate statistical models
6. Developing cost-benefit framework

Activities undertaken to deliver these steps include data collection and manipulation, data analysis, experimental programme, statistical model development and cost benefit. Following sections provide details on the study areas, and data collected for this study.

3.2.1. Study areas

The primary study areas for this research were subsystems from water supply networks of Severn Trent Water, UK. The entire STW network supplies water to a population of around 7.8 million, through a network of pipes with a total length of approximately 46,000km with around 24,000km of metallic pipes. The geographic area comprises of several counties, number of large cities as well as numerous rural and urban areas, with mixed domestic and industrial consumers. Historical developments of the area has lead to a combination of old and new pipes, with different pipe materials and installation methods across the network. Details on the distribution of pipes' characteristics are presented in chapter 4.

3.2.2. Model case studies

Three case studies are considered in this work, for model development and analysis, as described below.

Case study 1 (CS1) This is an area comprising of 39 DMAs, with total pipe length of 215.04 km. Total number of pipe failure records (outcome variables) in model development is 242. Further details on physical and hydraulic characteristics of this case study, and single-variate analysis of historical pipe failure records in this area are presented in chapter 4.

Case study 2 (CS2) A subset of the entire water network (CS3), the total pipe length in this case study is 1397.43 km, with 6774 pipe failure records. This case study covers the parts of the network for which high frequency pressure data was collected and used to estimate CPIS components (details of the sampling programme and estimation of CPIS components are explained in chapters 6 and 7, respectively).

Case study 3 (CS3) Entire network of the Company is considered as CS3, amounting to approximately 18093.95 km length of pipe, and 35729 number of pipe failure records. Further details on characteristics of the entire company's network is presented in chapter 4.

Table 3.1.: Characteristics of the case studies

Case study	Number of pipe segments	Total pipe length (km)	Number of pipe failure
CS1	3446	215.04	252
CS2	18502	1397.43	6774
CS3	268285	18093.95	35729

In this work, only metallic pipe segments and pipe failure records on metallic pipes are considered. Table 3.1 summarises the details of the case studies. Total number of pipe failures are the historical pipe failure records, between year 2003-2013.

3.3. Data

Analyses undertaken in this research used a range of variables and data. Table 3.2 lists the data used for each analysis.

Data used for various analyses in this research are taken from a variety of sources, some collected for and by the research programme, and others by extracting the required data from existing sources. Table 3.3 illustrates the data attributes that were used in this work. Source(s) of each data attribute, its general availability and steps and processes undertaken for preparing are also presented. Data sources referred to in this table include GIS (Geographic Information System), SCADA (Supervisory Control And Data Acquisition), SAP (Systemanalyse und Programmentwicklung - an enterprise database used at STW), Neasetbase (a Water Network Management System).

Table 3.2.: List of data attributes used in each analysis

Analysis	Data attribute used
Single-variate analysis	Pipe failure records, pipe age, material and diameter, surrounding soil corrosivity level and fracture potential indices, number of connected customers, number of valves on pipe
Cycle counting analysis	High resolution pressure data
Cumulative damage index development	Cycle counting data, steady state diurnal pressure variation, pressure variation
Hydraulic tracing analysis	Network physical and hydraulic characteristics, steady state pipe's flow level
Energy dissipation analysis	Output from hydraulic tracing analysis, pipe length, diameter and material, number of customers on pipes
Cumulative damage index at network nodes	Hydraulic tracing outputs, outcome of energy dissipation analysis
Logistic regression analysis	Pipe failure records, pipes age, material and diameter, surrounding soil corrosivity level and fracture potential indices, number of connected customers, number of valves on pipe, cumulative damage index
Cost-benefit framework	Total costs of pipe failure, total costs of intervention activities

Table 3.3.: Required data and general availability in STW

Data attribute	Sources	Availability	Preparation procedure
Network asset data	SAP, GIS	Good	Raw data sourcing, treatment of missing attributes, cross validation, spatial analysis
High frequency pressure data	Collected	Good	Pressure data collection programme
Pipe failure records	SAP	Average	Raw data sourcing and review for gaps and errors, data cleansing for erroneous data, missing data treatment, spatial analysis; also collected data via burst data collection application
Steady state pressure data	Hydraulic model	Good	Hydraulic model run (96-hour simulation time), script for multi-run 130 models, cross validations of data from GIS and hydraulic models
Network hydraulic details	Hydraulic model, network schematics, Netbase	Good	Hydraulic model runs, cross validation with Netbase data
Operational site details and working regime	SAP, SCADA, Netbase, site schematics	Average	Cross validation of data from multiple sources
Network topology	GIS, hydraulic model	Good	Cross validation of data (GIS, SAP)
Customer details and demand points	Billing database, SAP, GIS, Netbase	Good	Demand allocation and spatial analysis, customers points are associated to nearby pipes

Continued on the next page

Table 3.3 – *Continued from previous page*

Data attribute	Sources	Availability	Preparation procedure
Operational activities in the network	SAP	Poor	Separating operational and non-operational activities, Spatial and temporal analyses of data
Pipe failure modes and mechanisms	SAP, GIS	Poor	Cross validating data across sources, spatial analysis for precise burst location identification
Pipe surrounding soil records	Soil maps (Cranfield University, 1997)	Poor	Soil maps developed by the SSLRC (Cranfield University, 1997), high level soil type maps, not detailed for accurately defining pipe surrounding soil
Direct cost of pipe failure	SAP, Questionnaire	Average	Questionnaire data analysis, challenges in separating this from other related costs
Regulatory cost of pipe failure	SAP, Questionnaire	Good	Questionnaire data analysis, interruption data analysis
Indirect costs of pipe failure	Questionnaire	Poor	Most elements in this (details in section 10.2) not possible to obtain
Deterioration models characteristics	Existing model	Poor	Black-box approach for the Company's existing deterioration models

3.3.1. Missing data

Missing data in various datasets is a common challenge across the water utilities. There are several methods of dealing with missing data, such as:

- Likewise deletion: where any observation with one or more missing data

component is discarded

- Arithmetic mean imputation: where missing values are substituted by arithmetic mean value of existing observations. This method produces a complete dataset, however it causes reduction in the variables, resulting in biases (Wilks, 1932).

- Stochastic regression imputation: where regression equations are used to predict the missing values from the characteristics of the complete variables with addition of a residual term. This results in elimination of biases and restoring variability (Enders, 2010).

Within STW, a series of tools and methods have been developed that are used to populate missing data. For instance, there is a significant gap in available data about pipe's age. Based on the historical area development and other information regarding the area where the pipe was laid, age profiles are estimated. Such estimations are deemed to be the most reliable estimation of age and used in the current deterioration model in the Company.

In this work, a combination of existing company tools, spatial analysis and stochastic regression imputation is used to address missing data. Further details and examples of these are described in section 2.7.1 chapter 2.

3.4. Experimental programme

A sampling programme is designed in order to collect high resolution pressure data across the network. Design of the programme consists of studying physical and hydraulic characteristics of the network considering assets' operational performance, customer demands and other performance measures such as leakage, supply interruptions, etc. Details on the design of the experiment sample size calculation and sample distribution are presented in chapter 6.

3.4.1. Field work procedure

Figure 3.1 illustrate an overview of various steps in deployment and collection of high frequency pressure data.

After selection of the areas for logging (see chapter 6 for selection procedures), several datasets are sourced to review the physical and hydraulic characteristics of the network. This also includes consulting with local technical and operational team, in order to fully understand the network's hydraulics. A preliminary list of monitoring sites is developed, including locations within the distribution network and on operational sites (e.g. pumping stations).

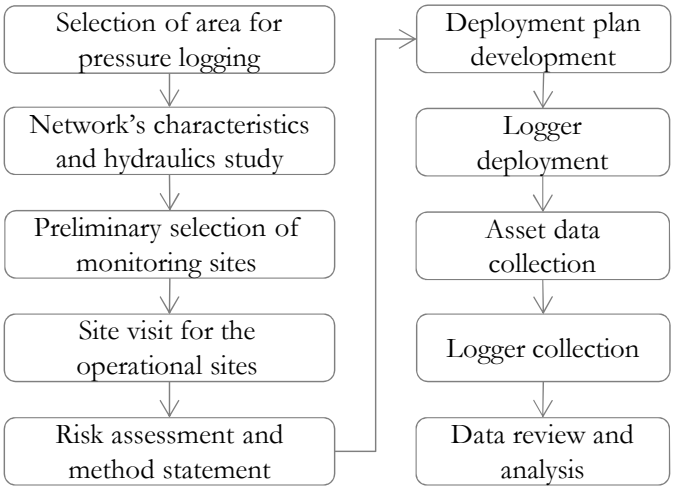


Figure 3.1.: Steps in collection of high frequency pressure data

Deployment of pressure transient loggers require tapping on the asset (e.g. pipes downstream of a pump), and it is often the case that such tappings either do not exist or they are not in working condition. A site visit is, therefore, arranged to ensure suitability and practicality of logging. Any site visit to an operational site requires risk assessment and method statement, to ensure safe delivery of the deployment activities. After the logistics for the deployment are in place, a development plan and map are prepared, identifying the exact locations of the logging points. Appropriate traffic management and other arrangements, as required, are undertaken in advance of the deployment visit. At the time of logger deployment, additional information regarding other operational assets, such as nearby flow meter, PRV characteristics, operation of reservoir valves, etc. are gathered. A copy of data collection form is provided in section A.1 in appendix A. Loggers are collected after required logging duration (details in chapter 6), followed by data download, review, and analysis.

3.4.2. Equipments and instrumentations

InfraSense, developed at Imperial College London, provided a sophisticated platform for capturing and analysing high frequency pressure data. The platform provided an easy to use deployment procedure and a robust data management and visualisation. The logging equipments, shown in Figure 3.2, comprises of a logger, and a detachable memory unit, battery module and transducer. These are designed for an easy *lift and shift* type installation, and the logger can be continuously in use (no down-time), moving from one site to another, attached with a new memory unit (and as required a fresh battery

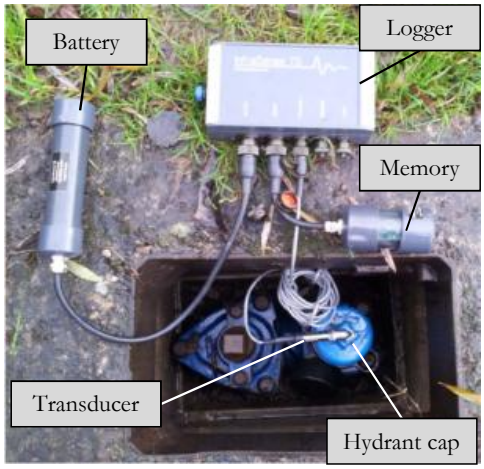


Figure 3.2.: InfraSense logging equipment, deployed on a fire hydrant

module). The logger can capture up to 128 pressure samples per second, with ±5 millisecond time synchronisation. The memory unit provides data capacity for continuous logging of up to approximately 4 months. Similarly, a fully charged battery module can support continuous logging for the similar period of time.

Collected data on the memory unit is downloaded by connecting the unit to the InfraSense server. The InfraSense platform provides a cloud-based system for storing the data and a robust Graphical User Interface (GUI) to visualise the data and locations of monitoring sites on a map. Example of the platform user interface is illustrated in Figure 3.3.

Low and high resolution recording Recent technological development in monitoring and data collection has enabled a significantly improved insight into the system behaviour. Figure 3.4 illustrate a comparison of data collected by a conventional 15-minute interval logger (here a Cello logger) against high frequency data collected by an InfraSense logger (here recording at 64 samples per second). The readings are for the same period of time at a fire hydrant (using a T-piece widget to enable multiple logging at the same location). While some conventional loggers can be configured to records at up to one-minute interval, as shown here, recordings form these loggers often fail to capture the rapid pressure changes.

3.4.3. Burst data collection

Data and information collected about mains repairs and bursts events are often limited and, when available, prone to inaccuracy. This is a significant but



Figure 3.3.: InfraSense platform GUI with plots of pressure data and monitoring site locations

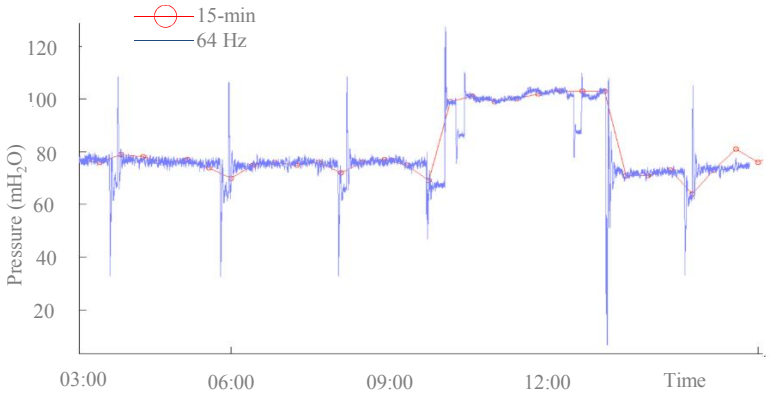


Figure 3.4.: Comparison of low (15 minutes interval) versus high (64Hz) resolution pressure data

common problem across most water utilities in the UK.

In order to tackle this issue, a programme of data collection from current main repairs across STW is developed. The collected data is utilised to develop a library of mains failure data and characteristics. Information collected include data about the burst and repair events, pipe and burst characteristics, environmental data and photographic survey. Table 3.4 lists various types of data and information collected.

To increase efficiency and accuracy of this exercise, and in order to facilitate integration of the collected data to other procedures and platforms (e.g. GIS), an electronic tablet-based solution is developed. A screen-capture of the tablet application developed is illustrated in Figure A.5, appendix A.

A pilot programme for implementation of the burst data collection exercise is undertaken. This included 15 repair teams over a period of four months. Examples from the collected data is illustrated in section A.2 appendix A.

3.4.4. Data analysis

The collected pressure data is analysed to extract individual cyclic loadings and associated event frequency. Cycle counting techniques are employed for this analysis. Cumulative pressure induced stress index is developed as a function of diurnal pressure variation and rapid pressure fluctuation (including pressure transient and pressure noises). Details of these analyses are described in chapter 7.

3.5. Statistical analysis

Single-variate analysis of historical pipe failure records, presented in chapter 4, is developed using a series of temporal and spatial analysis as well as frequency and probability distributions.

Logistic regression analysis is used to study multi-variate analysis of pipe failure records and factors contributing to the failure. Details of the method, model development and testing are described in chapter 8.

Most statistical analyses and model developments are undertaken using statistical software package R (RCoreTeam, 2014).

3.6. Cost-benefit framework

A cost-benefit framework is developed, as detailed in chapter 10, to evaluate cost-effectiveness of innovations and intervention measures aim at calming

Table 3.4.: Data collected when repairing mains failure

Category	Data/ information
Event data	1 Job number 2 Repair date and time 3 Manually recorded coordinates 4 Device recorded coordinates 5 Form filled by
Pipes details	6 Main ID 7 Mains material 8 Mains size 9 Mains cover 10 Mains external condition 10a Corrosion level 11 Mains external protection 12 Cathodic protection? 13 Joint type
Burst details	14 Burst location 15 Burst type 16 How was burst discovered?
Repairs details	17 Repair type 17a New mains material 17b New mains length 18 Sample tag and storage details
Surrounding ground	19 Ground surrounding main 20 Bedding under main 21 Surface use 22 Services close to the main 22a Distance of closest services to the main 23 Weather condition on burst (or repair) day 24 Weather temperature on burst (or repair) day
Additional information	25 Have there been bursts at or close to this burst over last few months? 26 Additional information as required
Photographic survey	27 Repair site photo 28 Leaking main photo 29 Mains external condition photo 30 Mains internal condition photo 31 Mains crack or holes photo 32 Repaired main photo

networks. In developing this, firstly, total costs of pipe failures as well as costs of intervention measures are estimated. Total cost of pipe failures often consist of several direct and indirect factors, from actual cost of repairing a burst pipe to various societal and environmental costs and regulatory penalties. A combination of expert interviews, questionnaires, business plan and financial reports of water utilities (with focus on Severn Trent Water financial data) is used to estimate total costs of pipe failures and relevant investments. Similarly, key costs of intervention measures are estimated, based on quantitative and qualitative components. A net present value technique combined with Monte Carlo simulations are employed to develop probability of the investment success of the intervention measure. List of data, methodologies and techniques and development of the cost-benefit framework is provided in chapter 10.

4. Single-variate analysis of historical pipe failures

Factors accelerating pipes degradation and failure were described in chapter 2. In this chapter, associations between these factors and historical pipe failure records are statistically investigated, by undertaking a series of single-variate analyses on pipe failures and the predictor variables.

Two case studies, Case Study number 1 (CS1) and Case Study number 3 (CS3), are analysed here. Overview description of these is provided in section 3.2.2, chapter 3, with further details provided below. Also, discussions on these case studies, and development of multi-variate models are provided in chapter 8.

In this chapter, firstly, physical and hydraulic characteristics of each area is reviewed. After that, single-variate analysis of pipe failure records and each predictor variable is investigated and discussed. The objective of this chapter is to assess the association between each factor, when taken individually, and the pipe failure.

4.1. Company network analysis

4.1.1. Network characteristics-CS3

The company's network comprises of approximately 46,000km length of pipes (around 24,000km are metallic pipes), supplying water to a population of approximately 7.8 million. For the purpose of hydraulic management, the network is divided to around 3200 DMAs and 130 control groups. A control group is a set of several hydraulically related DMAs grouped together, with the same supply source(s) and interruption contingency plans. Historical pipe failure records used for analysis in this work are from a 10-year period (2003-2013), as illustrated in Figure 4.1, summing to a total of approximately 77,000 pipe failure records, with around 57,000 on metallic pipes .

As illustrated in Figure 4.2a, metallic pipes accounted for approximately 50% of total mains length (around 18,000km of CI pipes, 5000km of DI pipes, with other metallic pipes comprising less than 0.5% of total pips length). CI

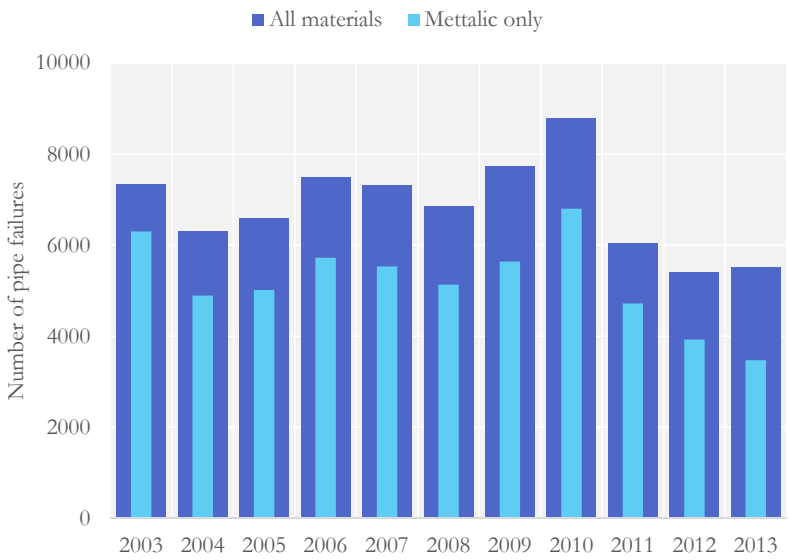


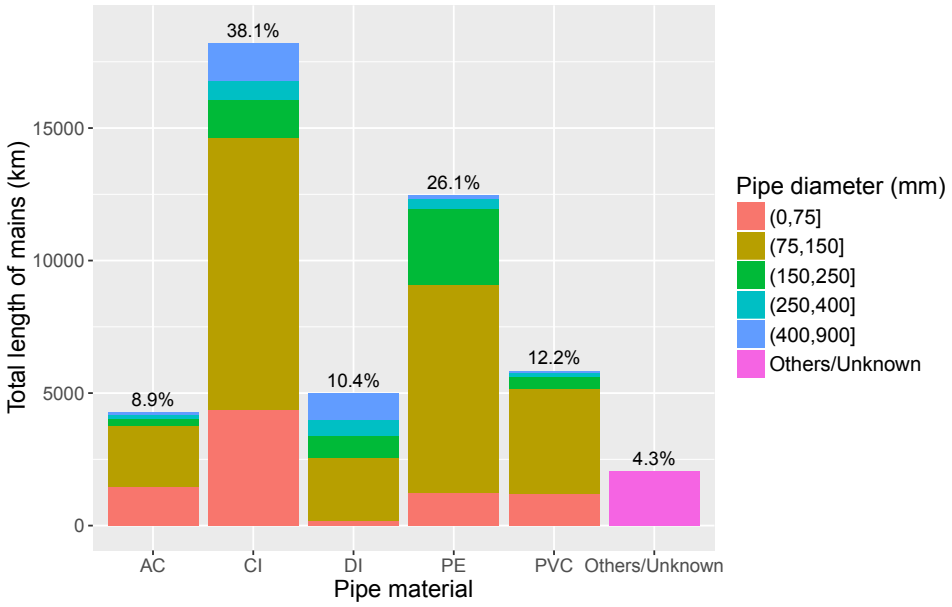
Figure 4.1.: Annual distribution of burst events

pipes are well represented across all pipe diameter sizes, dominant in all pipe diameter categories, expect for (150,250]mm range, where PE is the leading material. Diameter sizes for around 4% of pipe lengths are unknown, possibly suggesting these to be old CI pipes too.

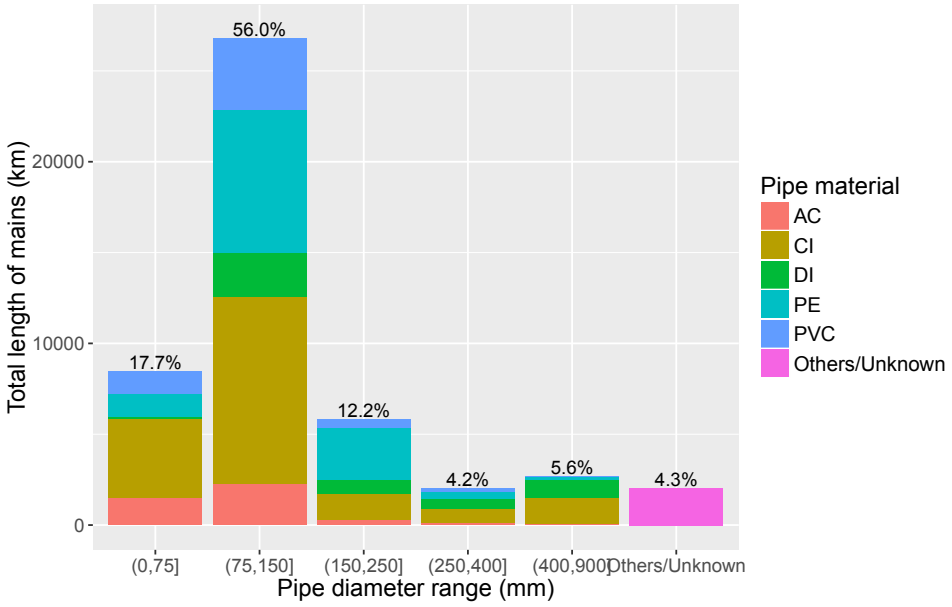
Figure 4.2b shows distribution of pipe diameter sizes across the network. Approximately, 85% of pipes are <250mm in diameter, with around 56% between 75-150mm diameter. In small diameter mains, PE pipes are closely following the leading CI pipes, although metallic pipes (CI and DI) are the dominant materials in larger pipe sizes (>250mm)

A large proportion of recent pipe installation are plastic pipes (PE and PVC), as well as DI (Figure 4.3a). The primary pipe material up to around 30 years ago had been CI, with pipes older than 90 years are almost entirely CI.

Up to around 30 years ago, proportion of various pipe diameter sizes over the years had been fairly unchanged, as shown in Figure 4.3b. Pipe diameters in the range of $(75, 150]$mm are the most used ones, followed by pipe diameters of $(0, 75]$mm. Other pipe diameters form relatively similar proportions at various time periods.

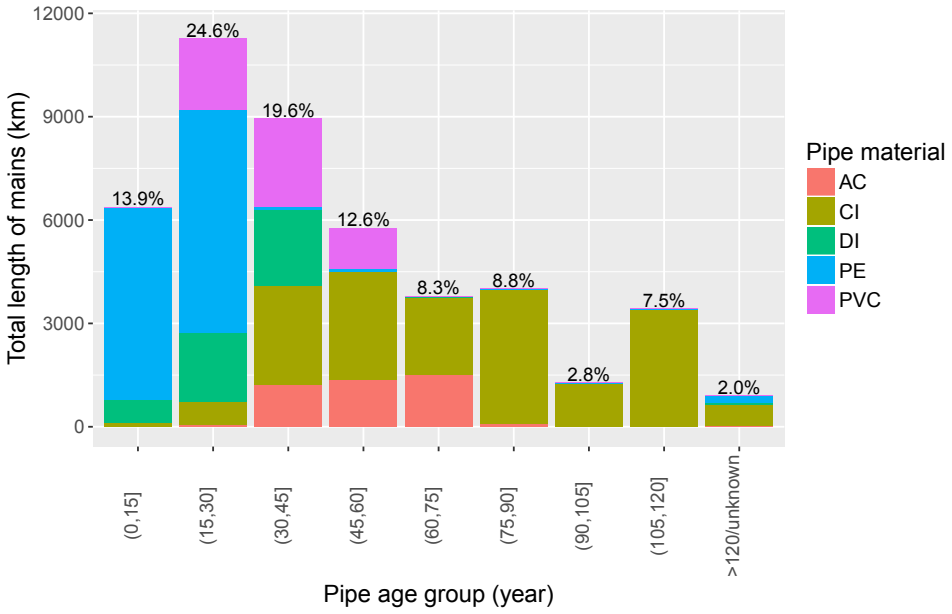


(a) Pipe material (and diameter)

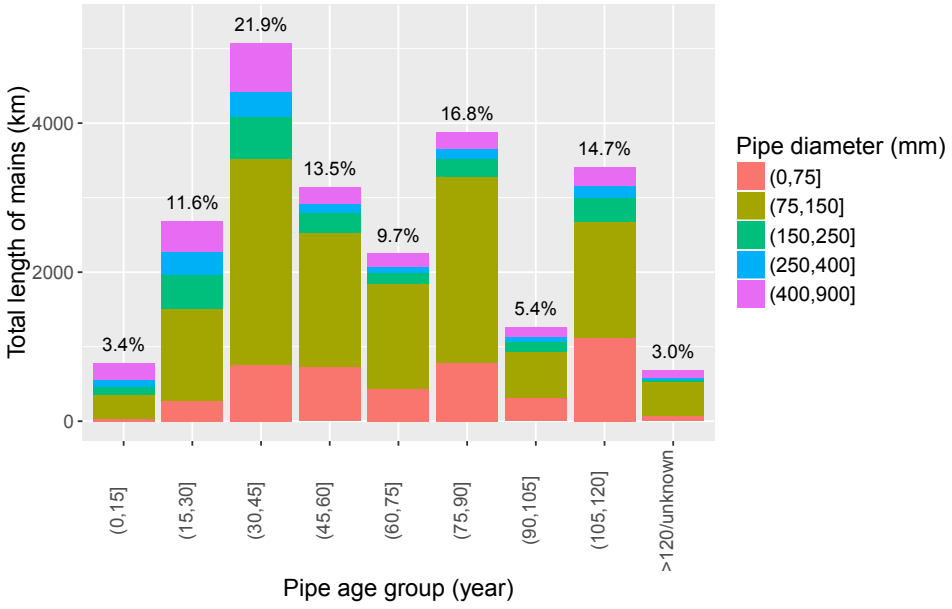


(b) Pipe diameter (and material)

Figure 4.2.: Distribution of pipe material and diameter



(a) Pipe age (and material)



(b) Metallic pipes' age (and diameter)

Figure 4.3.: Distribution of pipe age with material and diameter

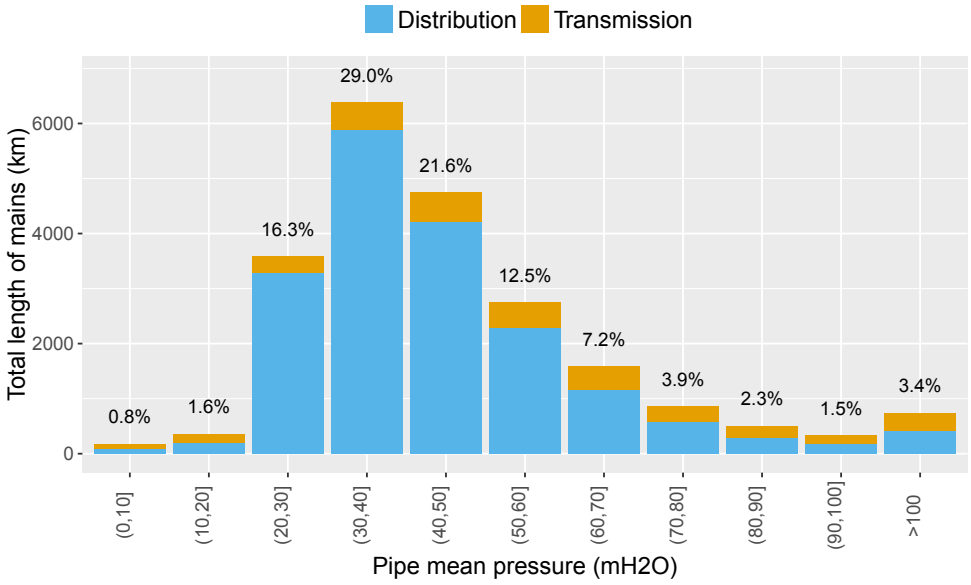


Figure 4.4.: Distribution of pipes' diurnal mean pressure for metallic distribution ($\leq$ 250mm) and transmission ($>$ 250mm) pipes

Figure 4.4 illustrates that diurnal mean pressure for a large proportion of the pipes is between 20-70mH_2O, with less than 15% outside this range. Similarly, shown in Figure 4.5, the diurnal maximum pressure is largely between 20-80mH_2O, with around 85% of pipes having diurnal pressure below 70mH_2O. Larger diurnal mean pressures ($>$70mH_2O) and maximum pressure ($>$80mH_2O) are mainly associated with transmission mains.

Diurnal mean and maximum pressures of less than 20mH_2O are possibly dead-ends, abandoned pipes not feeding any property or the pipes on operational sites.

Illustrated in Figure 4.6, steady state diurnal pressure variation in metallic pipes are shown to be largely below 20mH_2O with less than 15% of pipes experiencing pressure variations above this value.

Figure 4.7 shows the combined values of diurnal pressure maximum and pressure variations in metallic pipes. Most pipes with relatively low maximum pressure have a pressure variation of 10mH_2O or lower. Only a small proportion of these pipes experience large pressure variations. For pipes with maximum pressure above 50mH_2O, pressure variation is largely between 10-20mH_2O. As the pipe's maximum pressure increases, so is the proportion of larger pressure variation events.

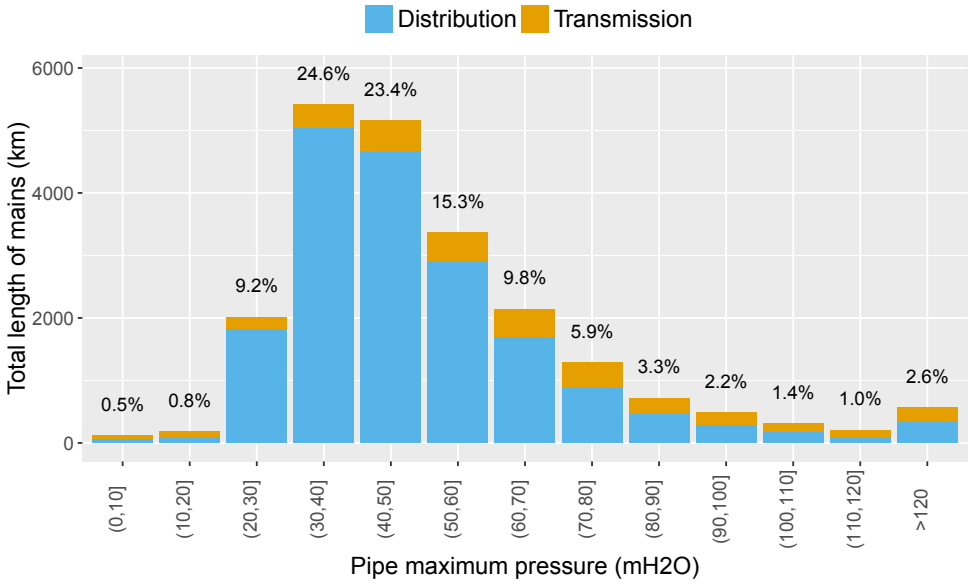


Figure 4.5.: Distribution of pipes' diurnal maximum pressure for metallic distribution ($\leq$ 250mm) and transmission ($>$ 250mm) pipes

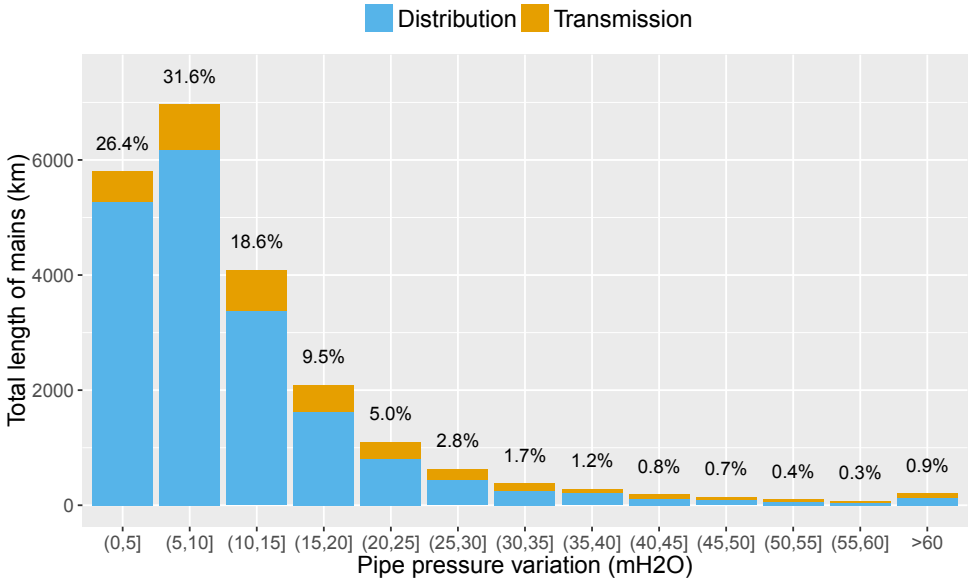


Figure 4.6.: Distribution of pipes' diurnal pressure variation for metallic distribution ($\leq$ 250mm) and transmission ($>$ 250mm) pipes

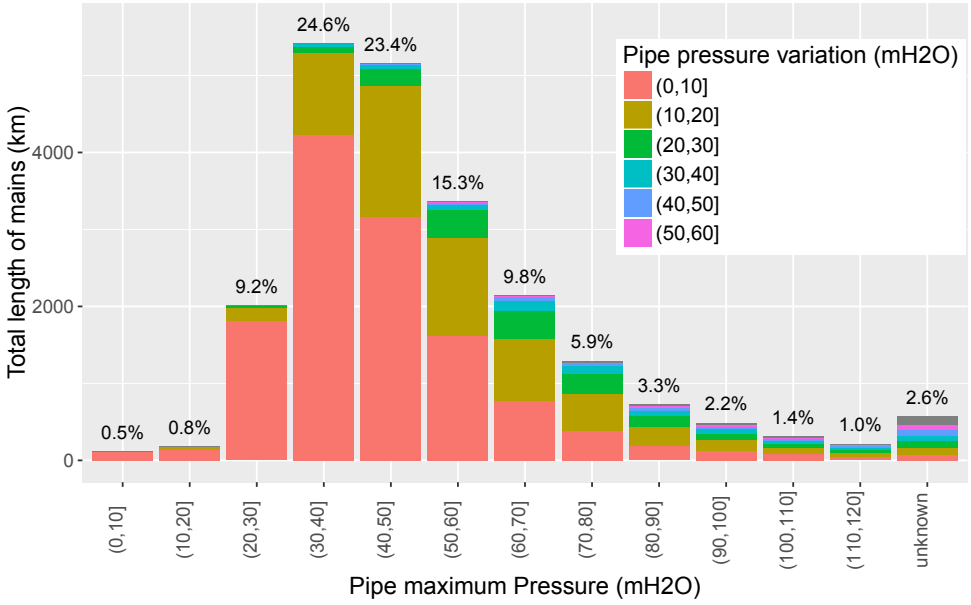


Figure 4.7.: Distribution of metallic pipes' maximum pressure and pressure variation

4.1.2. Single-variate analysis in CS3

There are several factors that can impact pipe structural degradation and failure. These factors can be categorised as pipe properties, external and internal loadings, and material deterioration. Pipe properties include pipe's material, length, size, structural properties, manufacturing and installation quality, pipe-soil interaction, previous break and maintenance history. Operational pressure, frost loads, traffic and third party loading are the main factors in the loading category. Material deterioration is mainly due to external and internal chemical, biochemical and electrochemical environment. Further details on factors impacting pipe degradation and failure are discussed in chapter 2.

Several analyses are undertaken to investigate the association between these factors and historical pipe failure records. Pipe failures are often result of the interactions between several factors. In the analysis in this chapter, relationships between the individual factors and historical pipe failure records are investigated. In chapters 8 and 9, combined impacts of these factors are studied in the multi-variate analyses.

An important factor in undertaking large scale statistical analysis, similar to the one performed here, is availability of required data. Table 4.1 summarise data availability for various pipe failure factors. For some of these, network-wide data is often not available. In other cases, a factor might act as a surrogate for other impacts. For instance, pipe age can be an indicator of the manufacturing

Table 4.1.: Pipe failure factors considered for single-variate analysis

Category	Burst factors	Data available	Surrogate factor/ analysis
Properties of pipe	Material	✓	
	Length	✓	
	Size	✓	
	Previous break	✓	
	Structural properties		Material
	Manufacturing and installation quality		Age
	Pipe-soil interaction		Soil
	Maintenance history		Age
Internal and external loadings	Operational pressure	✓	
	Loading from surrounding ground		Soil
	Frost loads		Soil, temperature
	Traffic loads		
	Third party impacts		
Material deterioration	External and internal chemical, biochemical and electrochemical environment		Soil, material

methods and installation procedure used during certain period of time. In such cases, surrogate factors or analysis are used to study the relationship with pipe failure.

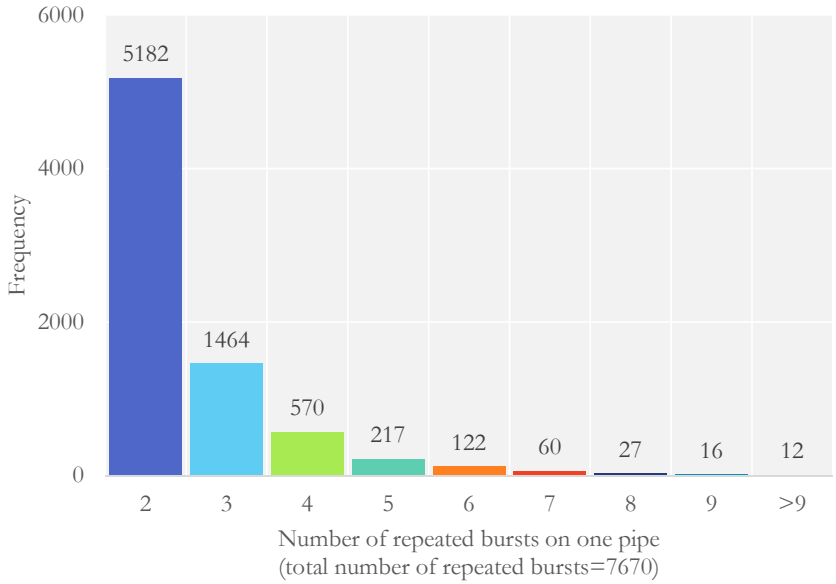


Figure 4.8.: Distribution of pipes with more than one historical bursts

Repeated pipe failure Majority of burst events have occurred on pipe with no history of failure, and approximately 15% of bursts (7670 out of approximately 57,000 burst records on metallic pipes) are repeated events (Figure 4.8). Approximately 60% of all subsequent pipe failures occurred within three months of a previous burst. This suggests an association between the activities undertaken when repairing broken pipes and future pipe bursts. Ground movements caused by excavations, backfilling and heavy vehicle movements, pressure transient events generated from valve operations, at repair and following pipe refilling, and third party impacts are possible factors that can impose additional stresses on degraded pipes causing failures.

Pipe material Analysing distribution of historical pipe failure across various pipe material illustrates that approximately, two third of mains failures have occurred on CI pipes, Figure 4.9a, while, as per Figure 4.2a, only around 40% of the of pipe length are CI. This makes CI as the leading pipe material in failure rate (Figure 4.9b), followed by AC, PVC, DI and PE. Combined with impacts from higher in-service age, poor manufacturing techniques and installation methods, physical characteristics of CI (impacts from corrosion, brittleness, fatigue failure, etc.) are likely to be the main reason for this.

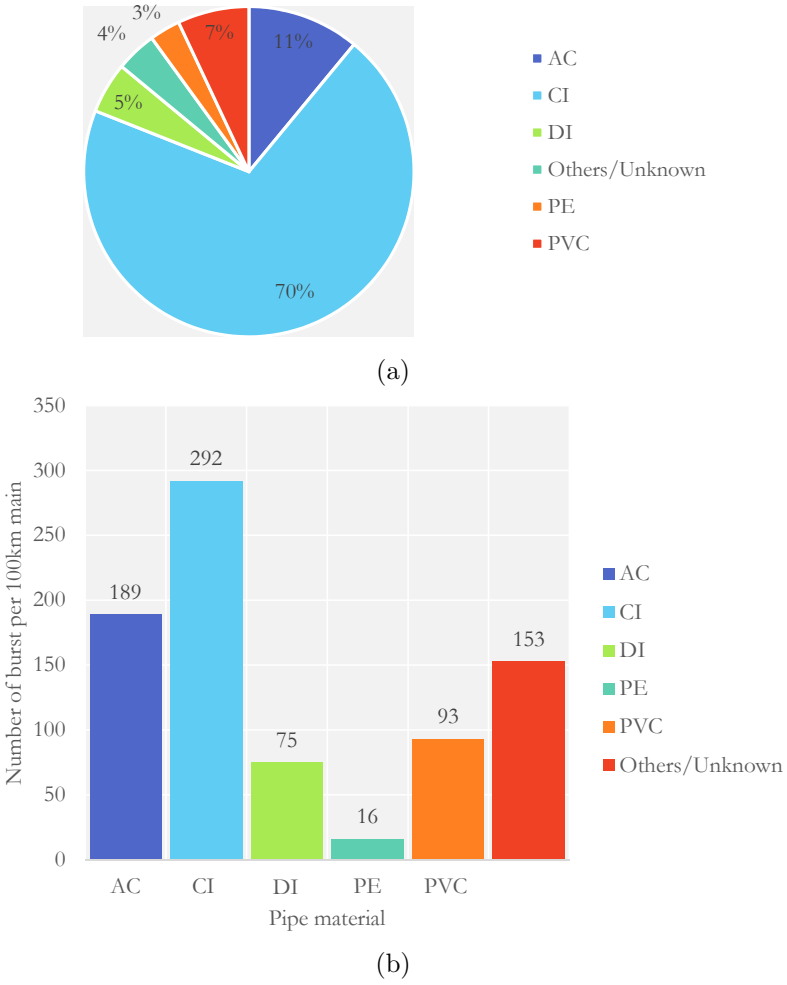


Figure 4.9.: Distribution of burst events across mains material (a) number of bursts (b) normalised for mains length

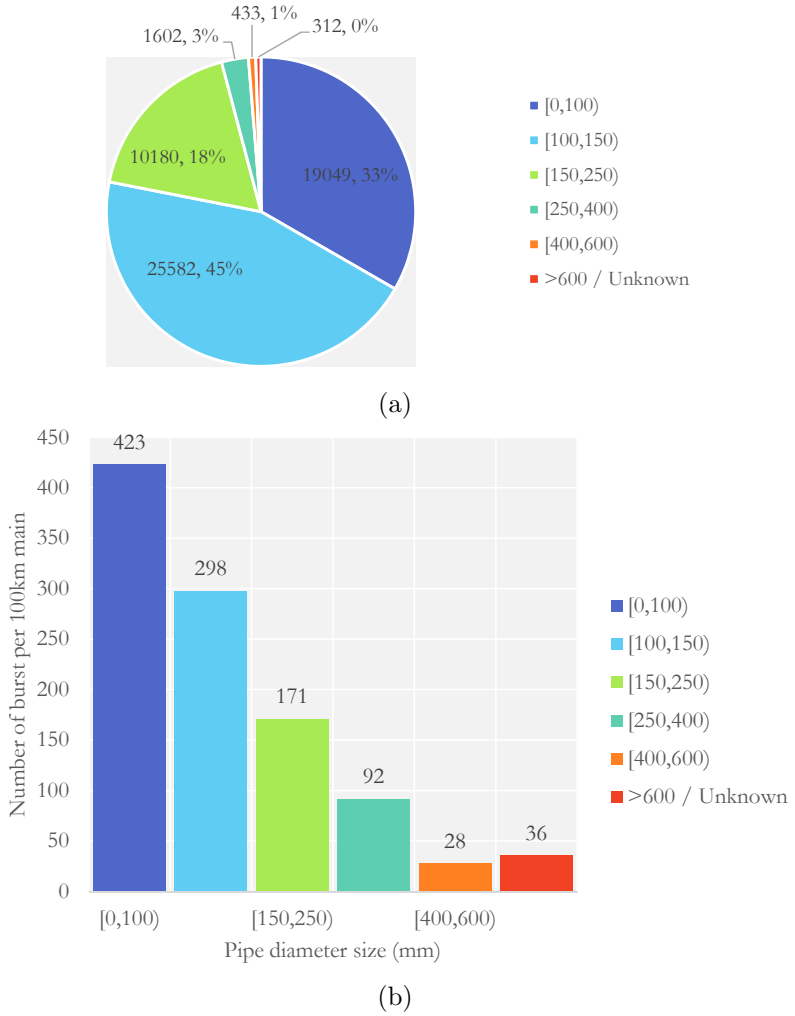


Figure 4.10.: Burst events distribution across mains diameters (metallic) (a) number of bursts (b) normalised for mains length

Pipe diameter size For metallic and non-metallic pipes, a vast majority of historical pipe failures ($\sim 95\%$) have occurred on distribution mains (< 250mm diameter). Larger pipes have failed considerably less frequently, although often consequence of a failure at a large diameter pipe is significantly greater. For failure rate, as illustrated in Figure 4.10b, a negative correlation is observed between mains' internal diameter and rate of mains failure. Larger diameter pipes have greater wall thickness, resulting in enhanced performance against deterioration mechanisms, such as corrosions, and also greater structural strength against imposed internal and external loadings and hoop stress.

Pipe age As pipes age, a significant increase in the risk of failure is expected during wear-out phase of the bathtub curve (Figure 2.2, chapter 2). Older

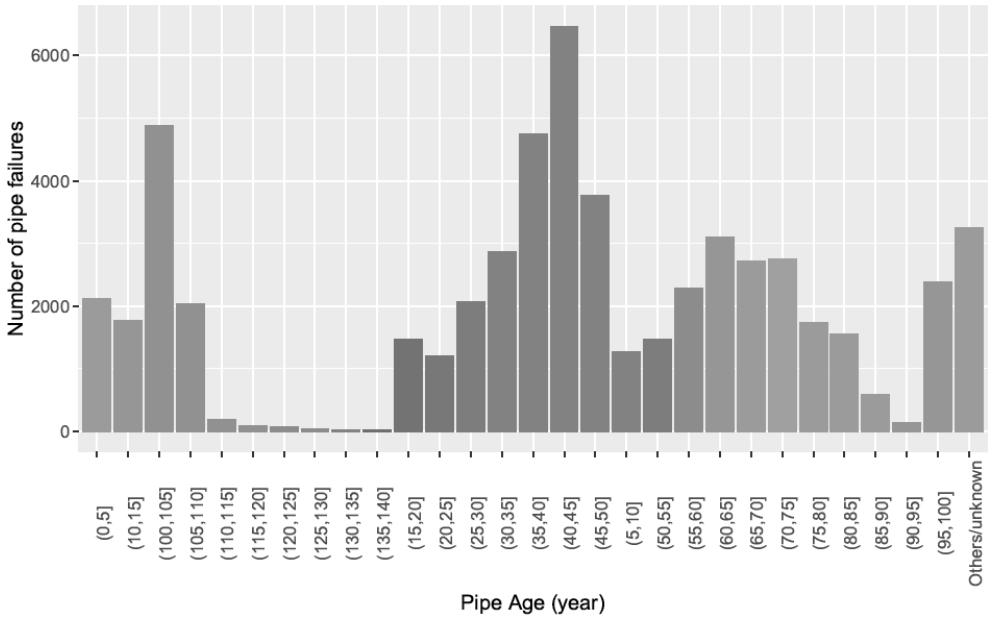


Figure 4.11.: Distribution of mains burst events for mains age (metallic)

pipes have been exposed to deterioration mechanisms for a longer period of time and have potentially developed flaws and weak points. Depending on the pipe material and surrounding environment, a proportion of pipe's wall thickness might have been lost and other structural degradation; that could increase probability of failure. However, the correlation between pipe age and risk of failure is not straightforward. Some previous studies, such as Boxall et al. (2007), have found a complex relationship between the pipes' failure and their age.

Distribution of pipe ages for metallic pipes for the 10-year failure records of this study, Figure 4.11, suggests no clear pattern in failure rate as function of age. Pipe age may act as a surrogate factor for several other factors, and also represent various historical developments. Factors, such as quality of laid pipes (manufacturing, construction and maintenance) and historical factors (e.g. World War periods) can add to the complexity of age-failure relationship.

Temperature Weather-related factors such as moisture intensity and duration of cold weather, rate of freezing, surface cover, mains depth, etc. affect the pipe failure rate. Nearly 60% of bursts occur in the Autumn and Winter seasons, with more failures in colder months of the year (Figure 4.12). There are, however, approximately 22,000 burst events (equivalent of 40% of total failures) occurring in the Spring and Summer seasons.

Pipe failure dates in the records used in this study are the dates associated

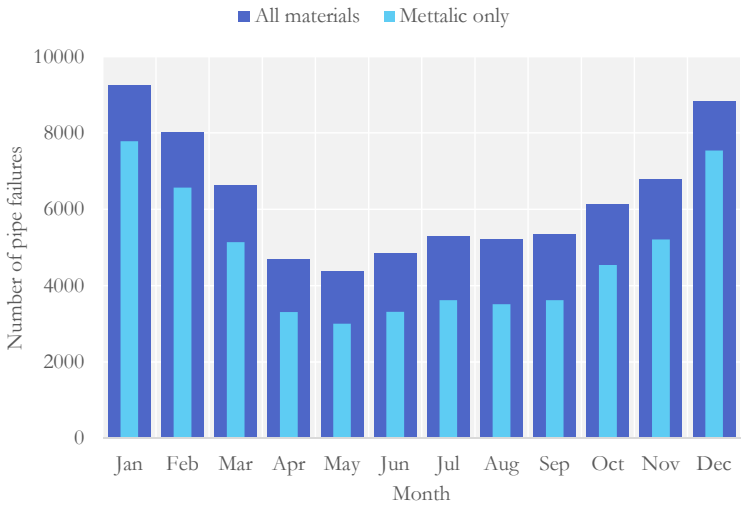
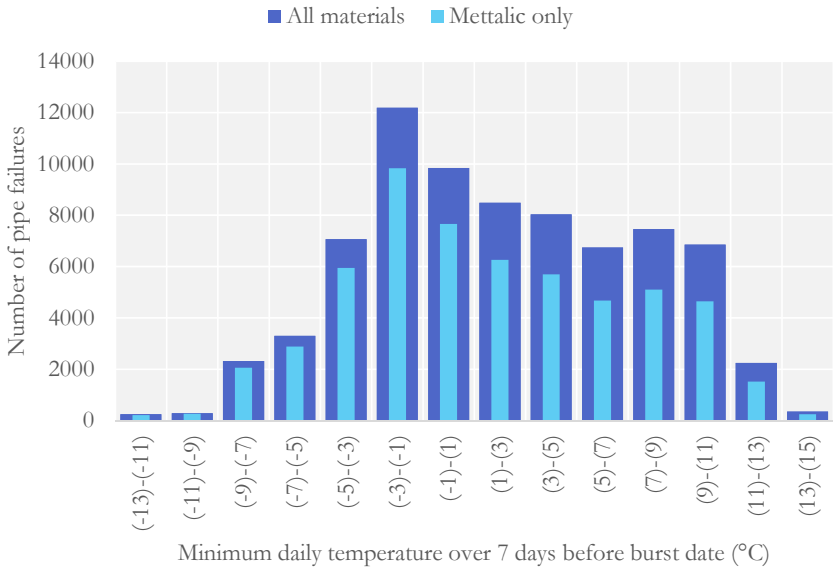


Figure 4.12.: Cumulative monthly distribution of burst events in metallic pipes

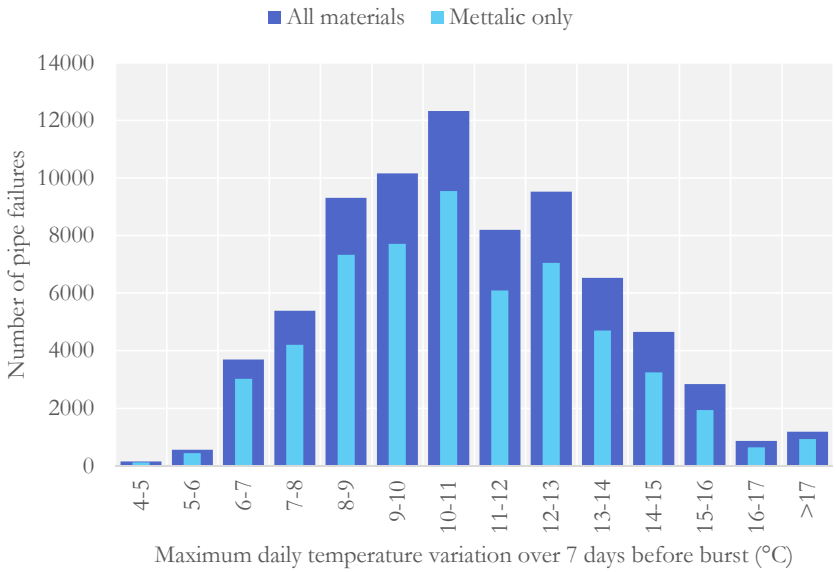
with the pipe repair activity. Depending on the failure type, the actual data of pipe burst might be different from the repair dates, potentially one or more days before the repair date. In studying association between ambient temperature and pipe failure, temperature values of the burst date and days preceding it are considered. Cold, particularly freezing, weather conditions can potentially contribute to increased rate of pipe failure. Distribution of pipe failures and the minimum daily temperature over the previous seven days leading to a pipe failure is presented in Figure 4.13a. The distribution is fairly similar for all pipe materials and metallic pipes only. The reduction in the absolute minimum temperature (even at freezing levels) do not show strong associations to increase in rate of failure.

Furthermore, significant change in the temperature can result in additional loadings from thawing and freezing the surrounding soil. Daily temperature changes within seven days preceding the failure date do not illustrate a positive correlation with the rate of pipe failure (Figure 4.13b). This distribution only considers the pipe failures where corresponding temperature had reached the freezing level (zero degree Celsius or lower).

Temperature is often considered as an important factor in pipe failure acceleration. The analysis undertaken here, investigating pipe failure associations with temperature as a single factor, does not provide sufficient evidence to support this. Inaccuracy of available data, varying pipe's depth of cover, characteristics of the surrounding soil, interaction between temperature and other factors, etc. are some of the factors that can affect the analysis.



(a)



(b)

Figure 4.13.: (a)Burst distribution versus minimum daily temp over 7 days before burst (b)Burst distribution versus maximum daily temperature change within 7 days preceding burst date

Table 4.2.: Combined soil characteristics versus historical burst distribution-metallic pipes

Combined soil characteristics	Number of bursts	Percent of total
FP: very low / non-aggressive	6761	20%
FP: low / slightly aggressive	2169	7%
FP: low / moderately aggressive	3187	10%
FP: moderate / moderately aggressive	6908	21%
FP: moderate / slightly to highly aggressive (sulphate)	3972	12%
FP: moderate / moderately to very highly aggressive (sulphate)	2217	7%
FP: high / slightly to highly aggressive (sulphate)	1778	5%
Other combinations	6013	18%

Pipe surrounding soil Adverse impacts of surrounding soils on development of pipe failure are often through stresses on the pipe as result of ground movement, as well as physical damages due to soil corrosivity. Soil characteristics representing these phenomena are often considered as fracture potential and soil corrosivity, grouped in several categories, as shown in Tables B.1 and B.2 in Sections B.1 and B.2, appendix B (Pritchard et al., 2013; Hembara and Andreikiv, 2012).

Soil maps developed by the Soil Survey and Land Research Centre (SSLRC) (Jarvis et al., 1997), are used to identify surrounding soil of the pipes and failure records.

As shown in Figure 4.14a, approximately 44% of the historical pipe failures occurred on pipes surrounded by soil with fracture potentials of *very low* to *low*, while a further 43% of pipe failures happened in soils with *moderate* level of fracture potentials. Pipe failure, normalised for mains length surrounded by each soil group, Figure 4.14b, shows a positive association between higher level of fracture potential and pipe failure records.

The associations between pipe failure and soil corrosivity is illustrated to be complex, Figure 4.15a and 4.15b. Although a positive association is observed in normalised rate of pipe failure as the soil corrosivity increases, records for group 4* and 5 show a drop in pipe failures in these soil categories. These are where soils are moderately to very highly corrosive, with, theoretically, higher risk of accelerating pipe degradation.

Table 4.2 present distribution of pipe failures versus several combination of soil fracture potential and corrosivity levels. In this table FP stands for soil fracture potential. The analysis illustrates that a large number of failures (8930 failures, approximately 27% of all failures) are on pipe surrounded by

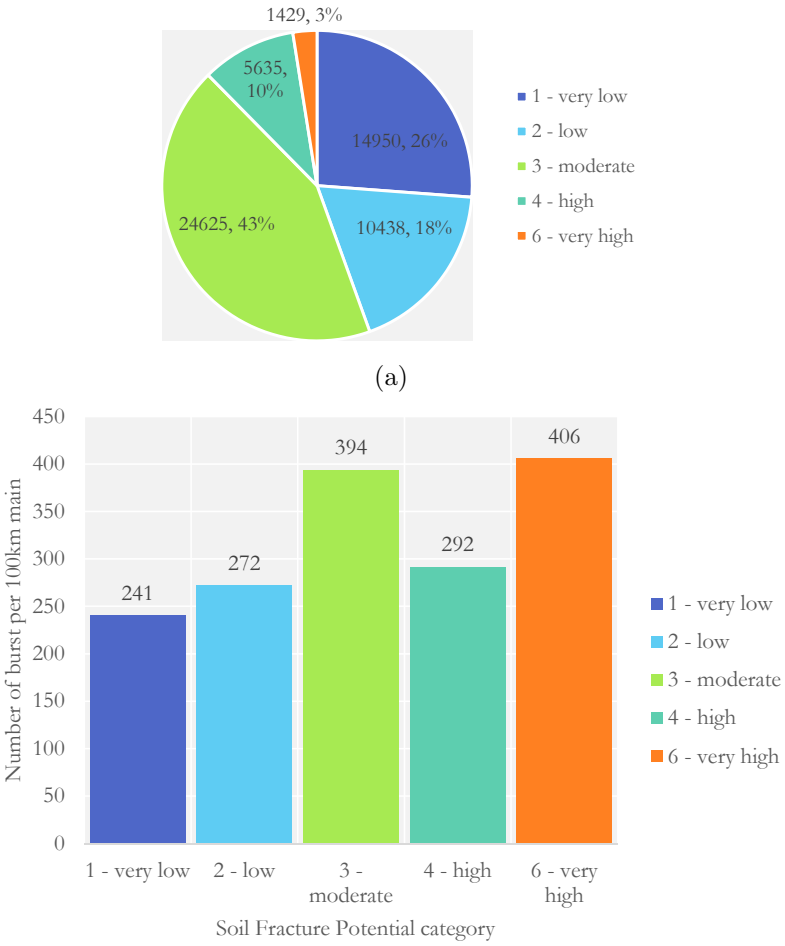


Figure 4.14.: Distribution of mains burst events versus soil fracture potential (a) number of bursts (b) normalised for mains length (metallic pipes)

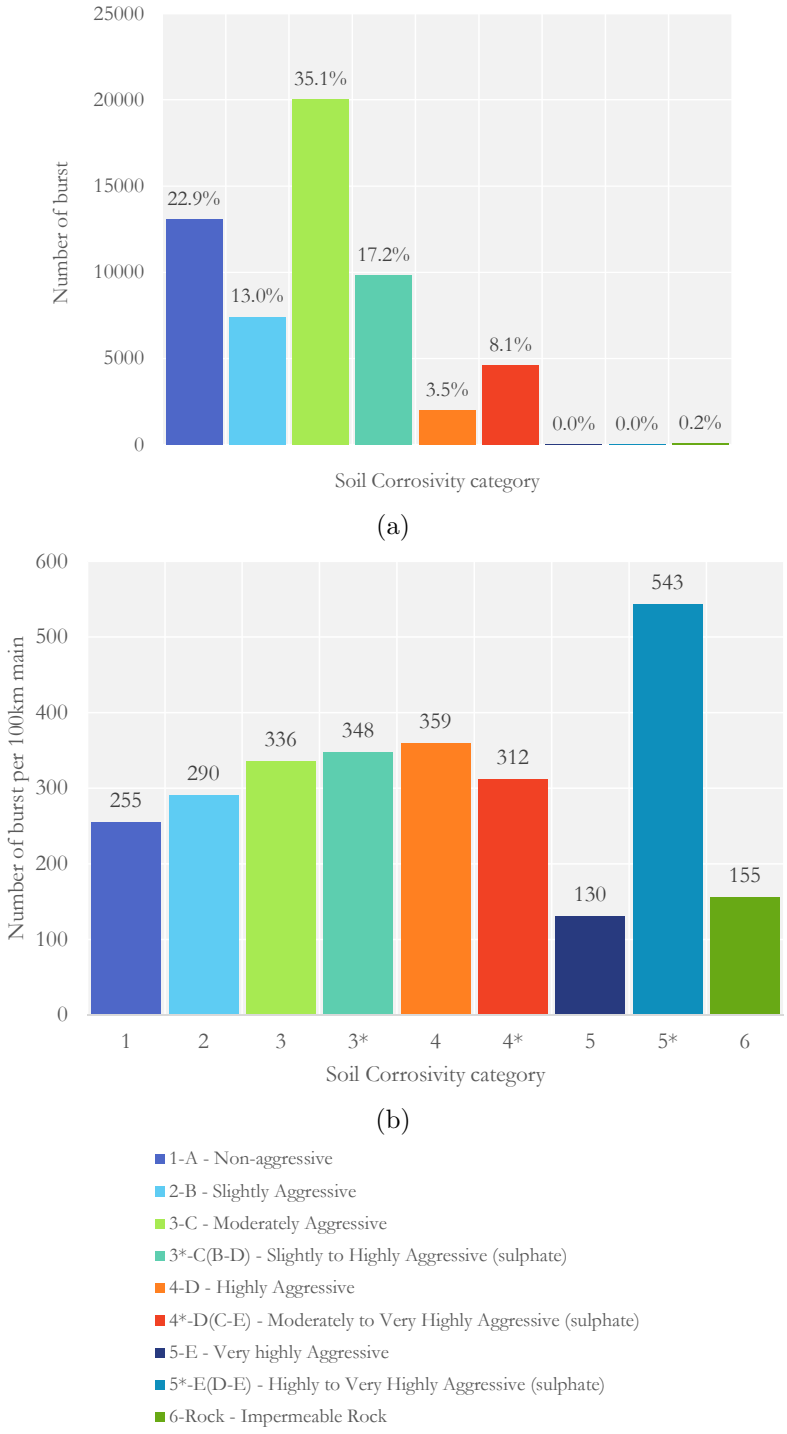


Figure 4.15.: Distribution of mains burst events versus soil corrosivity (a) number of bursts (b) normalised for mains length (metallic pipes)

non or slightly aggressive soil and where the fracture potential is very low or low.

Traffic loading Stresses imposed by traffic loadings could have adverse impact on pipe failure. Due to the nature of the analysis undertaken here (datasets from a geographically large area), reliable source of traffic data for the pipes with the historical failure records could not be obtained. Limited available data were found to contain significant errors and deemed unsuitable for further analysis. Therefore, no analysis could be undertaken for association of traffic loading and pipe failures.

4.2. Case study analysis

In this section, similar set of reviews and analyses are undertaken, this time on a subsystem of the network, a case study smaller in size. The dataset of this case study is used for the multi-variate analysis, detailed in chapter 8.

Below, an overview of the case study1 (CS1), its physical and hydraulic characteristics, and single-variate analysis of historical pipe failures are provided. The main criteria in selecting this case study were the following:

- Area characteristics (geographical location, area size, population and number of DMAs and hydraulic supply arrangements)
- Historical pipe failure records
- Presence of operational sites (large pump stations, reservoirs and automated valves)
- Availability of other data (pump operations statistics, burst investigation reports, etc.)
- Historical operational challenges (supply interruptions, discolouration incidents, etc.)
- Practicality of gathering high frequency pressures data
- Access to local knowledge and expertise

Schematic of this system, highlighting the locations of the DMAs, supply reservoirs and pump stations, etc. is presented in Figure 4.16a.

4.2.1. Network characteristics-CS1

This case study covers a hydraulically standalone area (with four transfer nodes connecting the system to the rest of the network). It comprises of 39 DMAs of varying sizes (from ∼50 to ∼3000 properties in each DMA). The total number of properties in the area is approximately 39,780, with a total pipe length equivalent of approximately $480km$, as shown in Figures 4.16b.

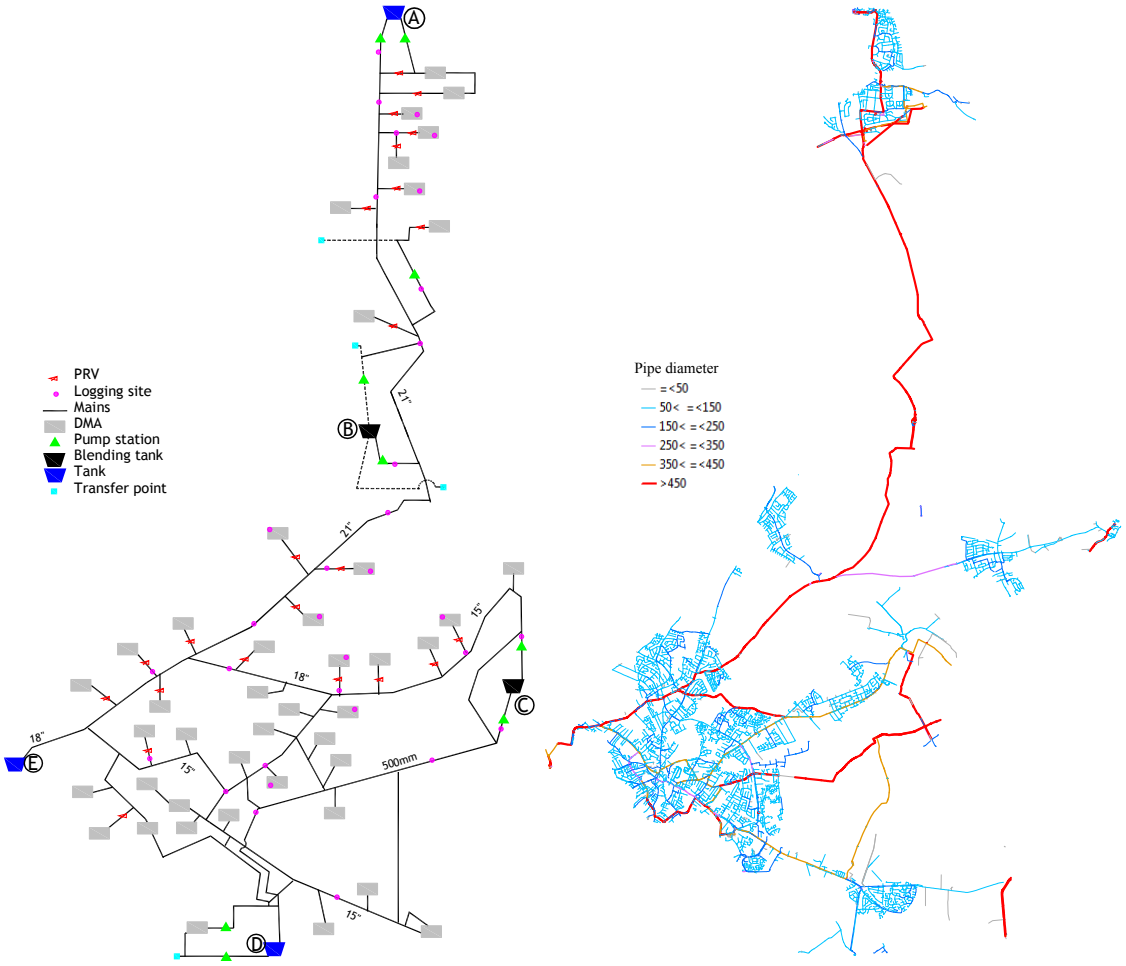
The highest and lowest property supplied in this area are at $137mAOD$ and $45mAOD$, respectively. The case study is spread over an approximate area of $375km^2$. Average zone night pressure (AZNP) of the area is approximately $41mH_2O$. AZNP is the property weighted average pressure in a zone measured at the time of the night line for leakage analysis purposes (usually between 3.00am and 4.00am) (MacKellar, 2003).

The area is supplied from three sources, labelled A, B and C in the system schematic, presented in Figure 4.16a. The total system demand is approximately 20 mega-litre per day. The combined operation of the booster pumps near the sources and the reservoirs (labelled as D and E in the schematic) are used to manage the demand and water pressure during normal and emergency operation of the system.

Figure 4.17 illustrates that the CI pipe material was the lead material up to around 50 years ago in this area. This is taken over by DI pipes for a couple of decades, before PE pipes dominated from around 30 years ago.

Recent decades have seen an increase in the proportion of larger diameter pipes being installed, as seen in Figure 4.18. There are far fewer <75mm pipes being installed. Although the overall level of new pipe installation is decreasing, the period between 30-45 years ago had seen a significant increase in new pipe installation, possibly due to specific developments in the area.

Cast iron pipes exist in all ranges of pipe diameter sizes including smaller than $75mm$ region, Figure 4.19a, while the majority of DI pipes are either smaller than $150mm$ or larger than $400mm$. As shown in Figure 4.19b, a large proportion of pipes are small diameter size, although, due to geographic characteristics of the area, approximately 15% of the pipes are larger than $250mm$ diameter size.



(a) (b)

Figure 4.16.: CS1 (a) system schematic (b) pipe diameter size distribution

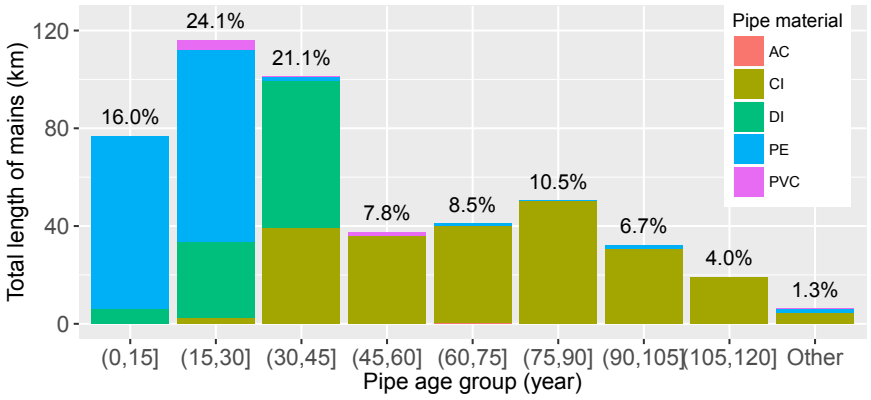


Figure 4.17.: Distribution of pipes' age and material in CS1

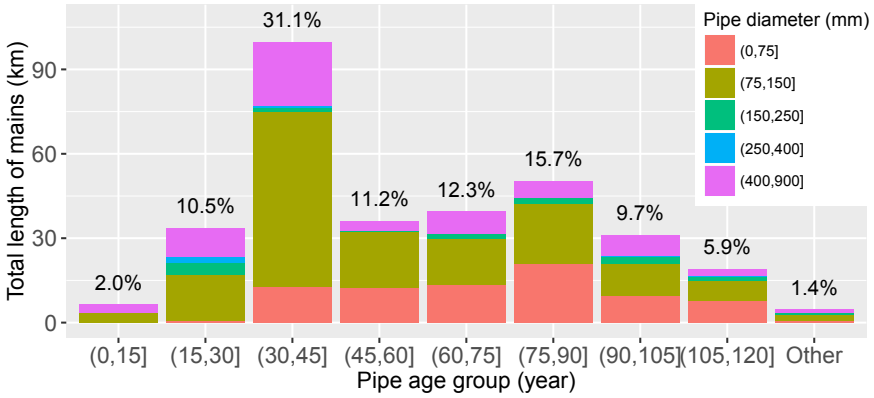


Figure 4.18.: Distribution of pipes' age and diameter in CS1 (metallic pipes)

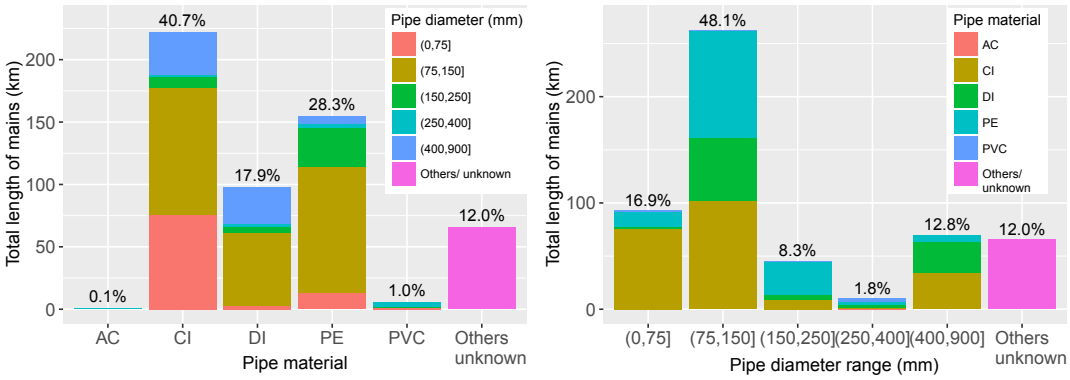


Figure 4.19.: Distribution of pipes' material and diameter in CS1

4.2.2. Single-variate analysis in CS1

Single-variate analysis of historical pipe failure records is undertaken, for a total of 409 metallic pipe failures, occurred between the years 2003-2013.

Annual distribution of historical pipe failure in this area does not follow a specific trend (Figure 4.20). This could be affect of environmental factors (particularly cold or wet annual climate condition) and maintenance regimes and level (impacted by the investment cycles).

The cumulative monthly distribution of pipe failure (Figures 4.21) shows a higher rate of pipe failure in colder months of the year, a trend observed in the larger dataset of the Company as well.

Per pipe length normalised distribution of pipes' ages at the time of failure, Figure 4.22, shows a complex association between pipe failure and its age. Similar complexity was observed in single-variate analysis of the Company's dataset.

As illustrated in Figure 4.23, majority of pipe failures in this area are on CI and DI pipes. When normalised per pipe length of each material, CI and DI pipes are still leading materials in the number of failure rates. Large figures for AC and PVC are due to the fact that there are very short lengths of AC and PVC pipes in this area, causing the normalised figures to escalate.

Figure 4.24 shows reduction in pipe failures as pipe diameter size increases. Similar trend is observed in the Company dataset.

The soil properties in this area is largely uniform, dominated mainly by soil in corrosion category of 1 (non-aggressive) and fracture potential of category 1(very low). Therefore, no particular insight can be gained from single-variate analysis of soil characteristics and pipe failures.

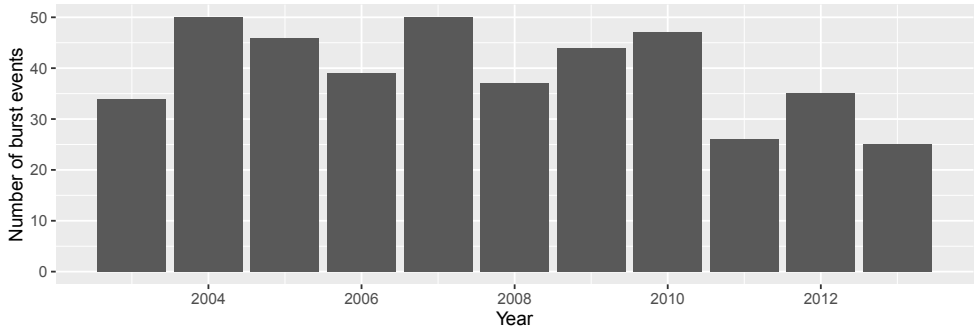


Figure 4.20.: Annual distribution of pipe failures in CS1

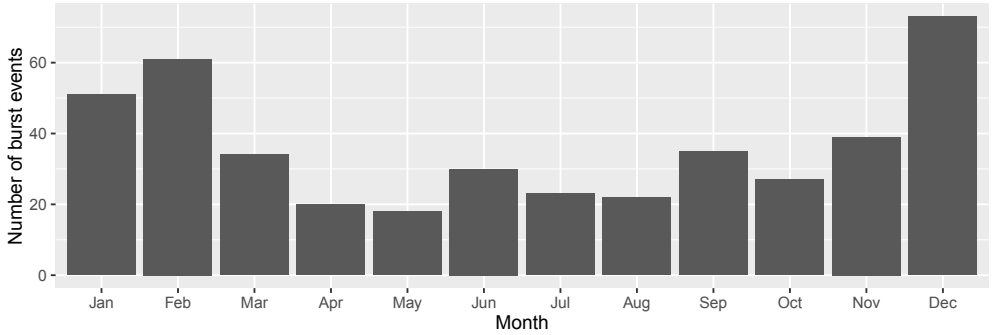


Figure 4.21.: Cumulative monthly distribution of pipe failures in CS1

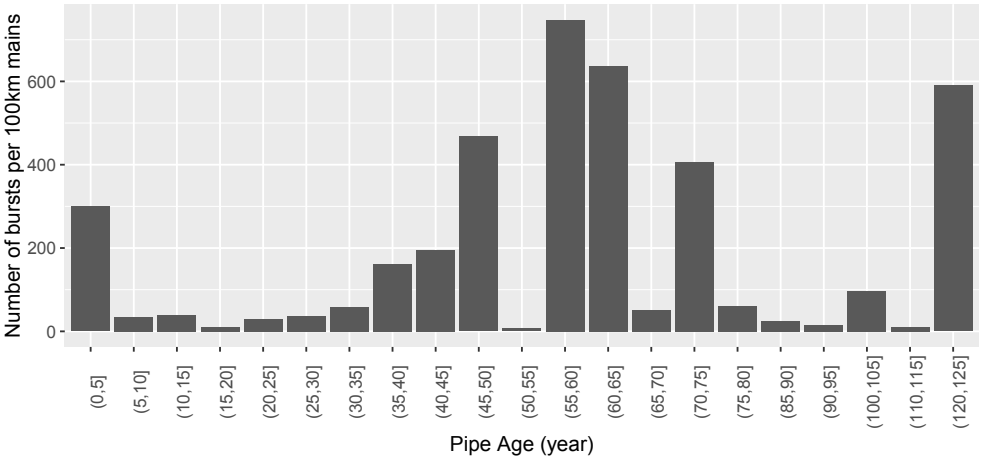
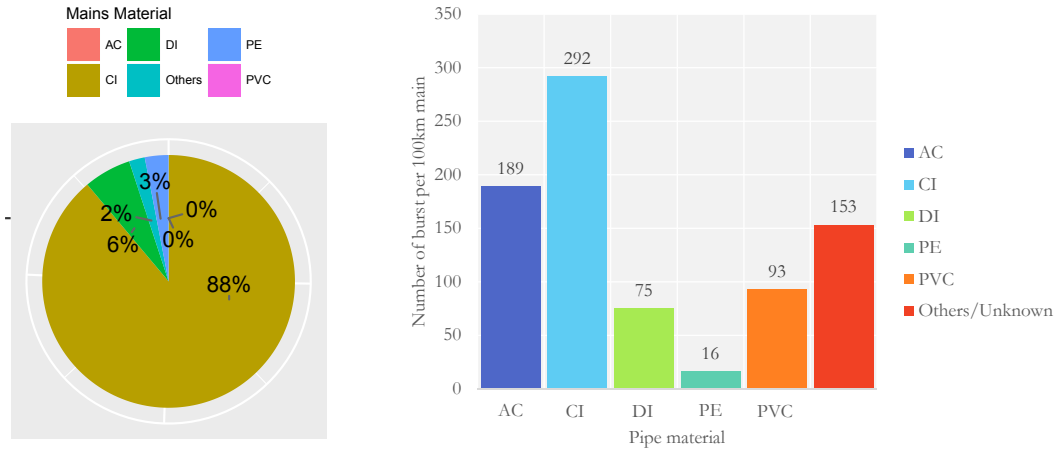
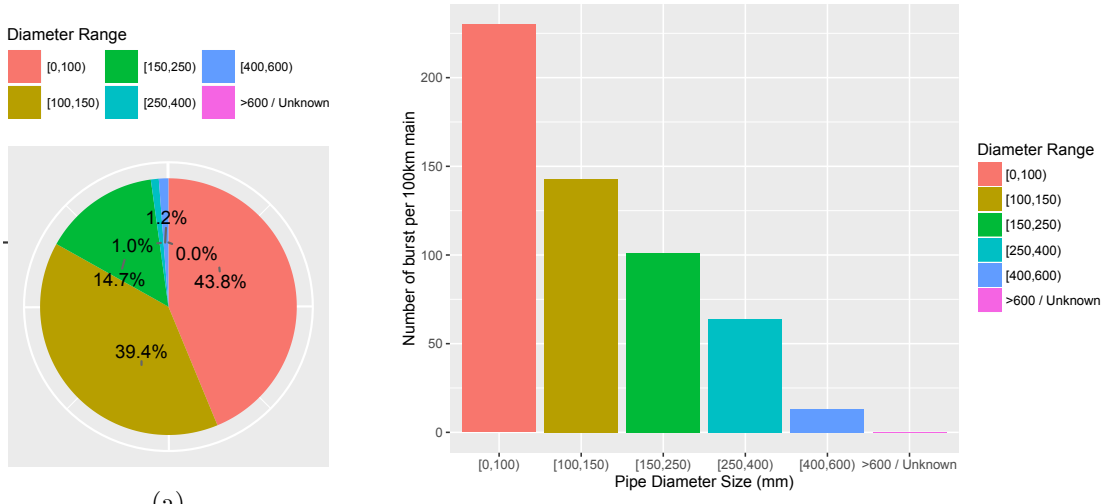


Figure 4.22.: Distribution of pipe failures for mains age, normalised for mains length in CS1



(a) (b)

Figure 4.23.: Distribution of failures in pipe material (a) absolute (b) normalised for mains length



(a) (b)

Figure 4.24.: Distribution of failures in pipe diameter (a) absolute (b) normalised for mains length

4.3. Summary and conclusion

Single-variate analysis can provide some insight into association between pipe failure and factors that can accelerate pipe deteriorations. Factors traditionally considered to have impact on pipe fail increase often fail to show a strong positive correlation with historical failure records. A large proportion of pipe failure records are found to be associated to the conditions that challenges the conventional view of pipe failure factors. For instance, a noticeable number of pipe failures have occurred in warmer months of the year or in surrounding soils that do not impose high level of corrosion and fracture potential to the pipes. Similar conclusions are observed from analyses on both the entire company's network (CS3) and the smaller subsystem case study (CS1). This challenges the widely accepted assumption that cold weather and corrosive soils are the (only) major contributors to pipe failures.

A National Sewers and Water Mains Failure Database, developed and maintained by UK Water Industry Research, UKWIR (Cima and Peters, 2009), hold a large volume of water mains failure records. The datasets are contributed by several (although not yet all) UK water utilities, providing a national picture on water and sewer mains failures. Similar observations were made when same analyses were performed on this larger database.

Pipe failure is a complex phenomenon and single-variate analysis might fail in capturing the interactions between various factors. Therefore, multi-variate analysis are undertaken to explore the associations further. The analyses in this chapter, however, suggest that there might be other variables that should be considered in the pipe failure analysis.

5. Fatigue failure of water mains

Stresses imposed by cyclic loadings in water networks can result in pipes' structural degradation and ultimate failures. Reviews of factors impacting pipe failure (chapter 2) and analysis of associations between pipe failures and these factors (chapter 4) have led to development of a new metric, to represent the hydraulic conditions of the network.

In this chapter, firstly, a brief background review of fatigue phenomenon and its primary characteristics are presented. Then, fatigue mechanism, including crack initiation, crack propagation and ultimate failure, are reviewed and the impact of mean stress is discussed. Fatigue analysis as well as primary factors impacting fatigue, such as material surface condition and environment, are then reviewed. Finally, a cumulative pressure-induced stress metric is introduced, that captures the combined effect of the system's steady, quasi-steady and unsteady pressure variations.

5.1. Fatigue failure

The impact of cyclic loadings from pressure fluctuations on the pipes is usually through mechanism of fatigue failure. Fatigue involves the initiation and evolution of cracks by loads that are too low to cause failure under monotonic loading (loads that are below the fracture strength of pipe material).

Shown in Figure 5.1, range of stress σ_r, stress amplitude σ_a, mean stress σ_m, maximum stress $\sigma_{\max}$, minimum stress $\sigma_{\min}$ and stress ration R, are defined as following:

$$\sigma_a = \frac{\sigma_r}{2} = \frac{\sigma_{\max} - \sigma_{\min}}{2} \tag{5.1}$$

$$\sigma_r = \sigma_{\max} - \sigma_{\min} \tag{5.2}$$

$$\sigma_m = \frac{\sigma_{\max} + \sigma_{\min}}{2} \tag{5.3}$$

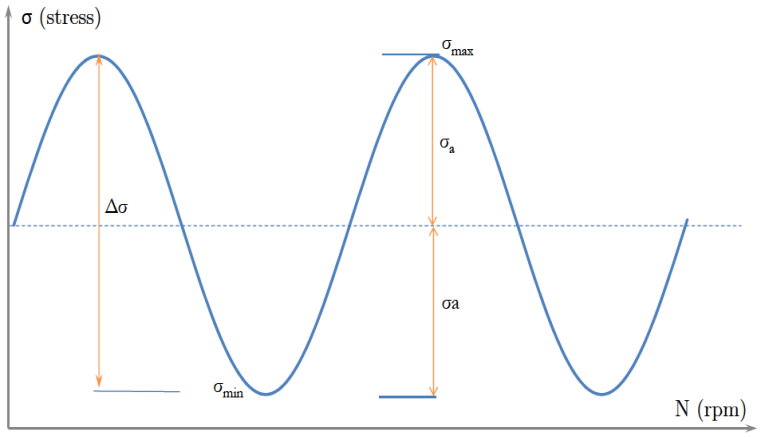


Figure 5.1.: Schematic of cyclic loading profile tension-tension cycle

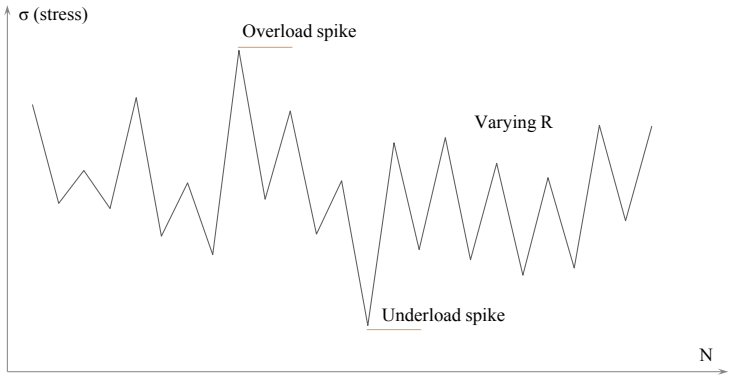


Figure 5.2.: An indicative unsteady pressure profile

$$R = \frac{\sigma_{\min}}{\sigma_{\max}} \tag{5.4}$$

The loading profile could be symmetric ($\sigma_m = 0$), asymmetric ($\sigma_m \neq 0$) or random. Loading profiles as a result of diurnal and seasonal pressure variations, in most cases, could be approximated to an asymmetric profile. As in most cases, pipelines are exposed to positive pressure, a tension-tension cycle (R>0) graph can represent the loadings imposed to the pipe. While pressure variation (daily and seasonal) can be represented by relatively constant stress relation, R, pressure transient events and hydraulic instabilities need to be approximated with a random profile (varying R) (Figure 5.2).

Considering the loadings, fatigues are generally governed by at least two parameters: stress (or strain) range and mean stress (or strain). The overall effect of stress (or strain) range is characterised by stress-life ($S - N$) curves, as expressed in the Figure 5.3. This consists of crack initiation and crack

propagation curves, the combined effect of these creating the total life curve.

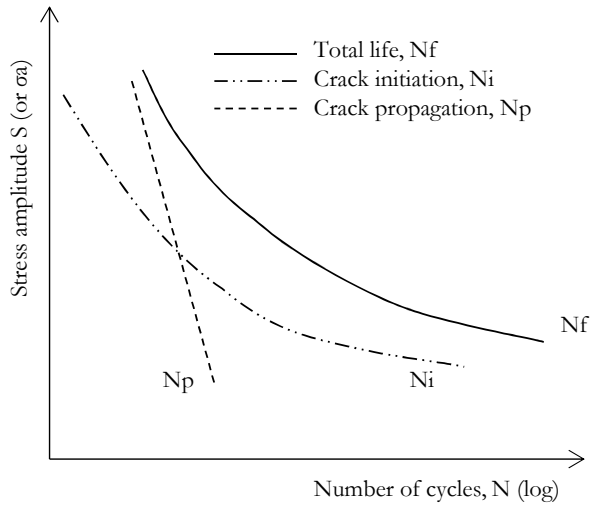


Figure 5.3.: Curve for total life, crack initiation and propagation

5.1.1. Fatigue mechanism

The fatigue mechanism consists of three steps: crack initiation and formation, crack propagation (the stable incremental enlargement of the crack) and rapid unstable fracture, i.e. failure.

Fatigue initiation

Initiation of fatigue occurs at microstructure-scale nucleation sites within the material. Such sites could be inclusions, pores, soft grain regions or micro-voids (Hertzberg, 1996). The presence of micro-scale stress concentrators, such as threads, grooves and notches, manufacturing flaws, pipe deformation, sites of corrosions pits and stress-corrosion cracks, etc. enhances this process.

In fatigue's '$S - N$' curve, when the stress level is low and the deformation is primarily elastic, the fatigue is known as *high cycle type*, i.e. fatigue-induced failure requires a large number of cycles. When the stress level is high enough for plastic deformation to occur, the fatigue is known as *low cycle type*, where low number of cycles are needed for fatigue failure. On the left of the curve, referred to as *low cycle fatigue*, a large amount of plasticity, developed at the stress concentrators, can hasten the fatigue initiation process. In the order of 10^5 or more cycles (N), the amplitude of stress imposed on the material will be fairly low.

When stress amplitude is sufficiently low (lower than S_e, 'endurance limit'

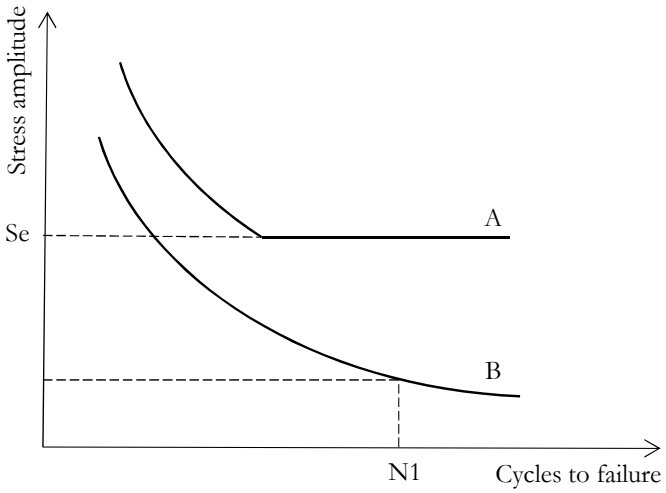


Figure 5.4.: $S - N$ curve for different materials

in curve A of Figure 5.4), the asset lifespan is infinite with regards to fatigue failure. Some materials do not exhibit an endurance limit, i.e. at any stress amplitude (even very low), if exposed to a sufficiently high number of cycles, the material will eventually fail under fatigue (curve B in Figure 5.4). For instance, steel's fatigue behaviour follows curve A, while aluminium's behaviour is similar to curve B. *As-finished* pipe and welds may not exhibit an endurance limit (Markl and Louisville, 1955). Endurance limit may also disappear (curves A to B) due to periodic overloads, corrosive environment (fatigue corrosion interaction) and high temperature (Engrasp, 2008). Other factors impacting this limit may be: surface finish, temperature, stress concentration, notch sensitivity, size, environment (vacuum, relative humidity, corrosion), creep and time-dependant phenomena, etc.

The $S - N$ curves are developed from tests to failure. Therefore, the number of cycles includes the propagation and final fracture phases of fatigue. In the absence of an initial crack, it is assumed that these phases comprise a very small proportion of the overall fatigue life (Rosenfeld and Kiefner, 2006).

It is recommended that the maximum stress applied by the service loading on the pipelines to be considered 90MPa (Marshall, 2001). This is where there is considerable ground movement and in cold weather. Figure 5.5 shows the critical defect size for various wall thickness iron pipes for this stress level, 90MPa, where fracture toughness varies with wall thickness as per equation below:

$$\sigma_f = \frac{K_c}{\sqrt{\pi a}} \tag{5.5}$$

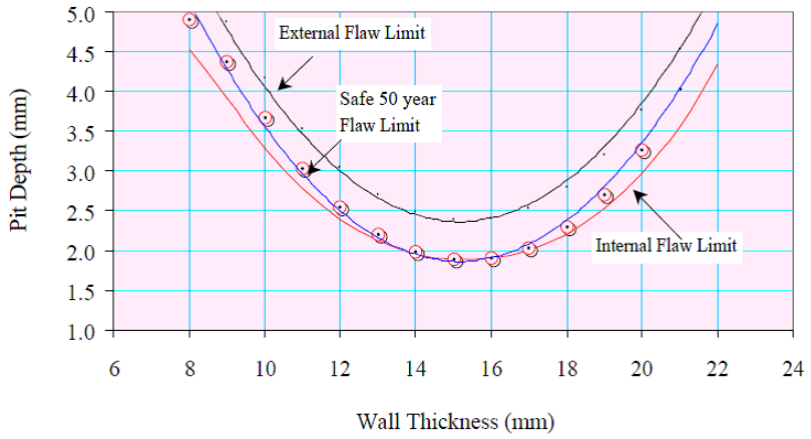


Figure 5.5.: Limits on critical pits to prevent crack growth with a stress of 90MPa-adopted from Marshall (2001)

Where σ_f is failure stress in MPa, K_c is fracture toughness (critical stress intensity) in $[MPa\sqrt{m}]$ and a is crack size in $[m]$. Fracture toughness is the material property describing the ability of a material containing a crack to resist fracture.

Crack initiation period is the amount of component life on crack nucleation and short crack growth. The component life spent during the long crack growth in the crack propagation period. Transition from the initiation to propagation is not very clear, but for steel, the size of crack at the end of initiation period, a_0, is typically range from 0.1 - 1mm (Lee et al., 2005). There are linear elastic fracture mechanics approaches to calculate a_0 for smooth specimens, but for water pipes with not-smooth surface, Dowling (1998) suggest 0.1 to 0.2 times the notch-tip radius for notched specimen.

With a sufficiently large number of cyclic loadings, fatigue can occur eventually even if the material surface is free of stress-concentrating features. This however requires a very large number of cycles that might not be the common scenario for water pipes.

Crack propagation

Propagation necessarily concerns a crack that already exists. Propagation of a fatigue crack in service is governed by Paris Law:

$$\frac{da}{dN} = C[\Delta K]^n \tag{5.6}$$

Where $\frac{da}{dN}$ is the increment of crack extension per load cycle, in $[mm/cycle]$ and ΔK, in $[MPa\sqrt{m}]$, is the amplitude of the range of alternating crack-tip

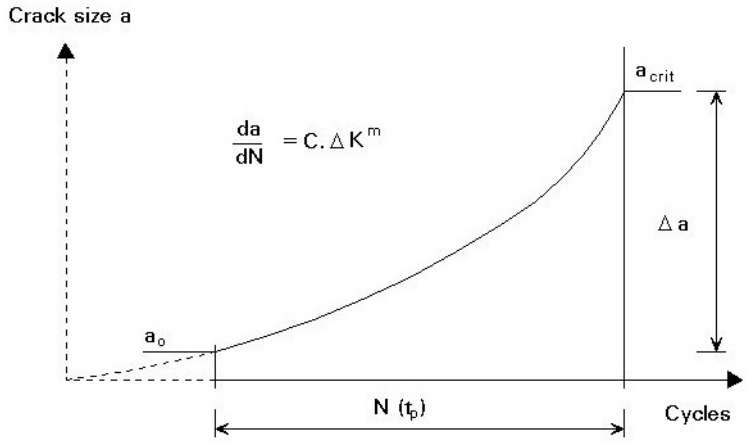


Figure 5.6.: Crack growth rate development with accumulated cycles

stress intensity associated with a given load cycle acting on the crack of size a, and C and n are material properties.

The crack-tip stress-intensity is an expression of the theoretical stress at the tip of a crack, derived from linear elastic fracture mechanics as:

$$\Delta K = f(geometry) \times \Delta\sigma \times (\pi a)^{1/2} \tag{5.7}$$

where $\Delta\sigma$ is a nominal applied stress (fluctuating over a range), a is the crack size in $[m]$, and ΔK is expressed in $[MPa\sqrt{m}]$.

Values for C and n are influenced by load cycle frequency, stress ration, and also chemistry of the environment the crack is exposed to.

The crack growth rate increases exponentially, as per Paris Law, as load cycles accumulate, Figure 5.6. This means that a small crack may remain small for a long time, and suddenly expand to a failure limit. Effect of larger initial flaw is to speed up the crack development, causing failure at much reduced load cycles.

In Figure 5.7, regime I (threshold) represents the area where fatigue cracks are not propagating. In regime II, representing Paris Law, there is a linear relationship between the fatigue crack growth rate ($\frac{da}{dN}$) and stress intensity factor range (ΔK). Regime III is where unstable crack growth develops result in fracture.

Final fracture and effect of mean stress

When the crack growth rate accelerates and crack size has reached a critical size, the pipe could fail at the next applied load cycle. Such critical size depends on material strength, nominal stress and fracture toughness.

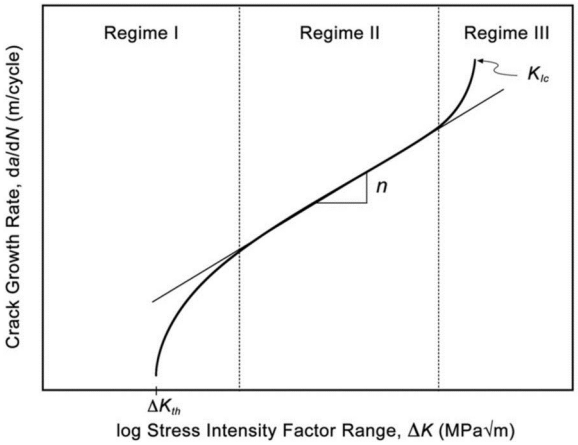


Figure 5.7.: Stress intensity factor vs. crack growth rate

To investigate the effect of mean stress, the 'Haigh diagram' can be developed, an example is shown in the Figure 5.8. A substantial amount of testing is needed to develop such a graph, as various empirical relationships are developed relating the alternating stress to mean stress. These methods connect the endurance limit on the alternating stress to either the yield strength, S_y, ultimate strength S_u, or true fracture stress S_f on the mean stress axis. These include Goodman line, Gerber parabola, Soderberg line as well as Morrow line. Goodman is often used due to mathematical simplicity and slightly conservative value.

5.1.2. Fatigue analysis and characteristics

Cumulative damage models

The effect of fatigue is often assumed to be cumulative (Lee et al., 2005). There are various approaches to develop the cumulative damage models. This could be a linear damage model, double linear damage rule, damage curve approach, inverse power law model, etc.

Miner's rule states that when there are k different stress levels and the average number of cycles to failure at the i^{th} stress, σ_{a_i}, is N_i, then the damage fraction, C, is:

$$\sum_{i=1}^{k} \frac{n_i}{N_i} = C \tag{5.8}$$

Where n_i is the number of cycles accumulated at stress σ_{a_i}; C is the fraction of life consumed by exposure to the cycles at the different stress levels. To quantify the damage, W_i is developed by the product of stress and its number

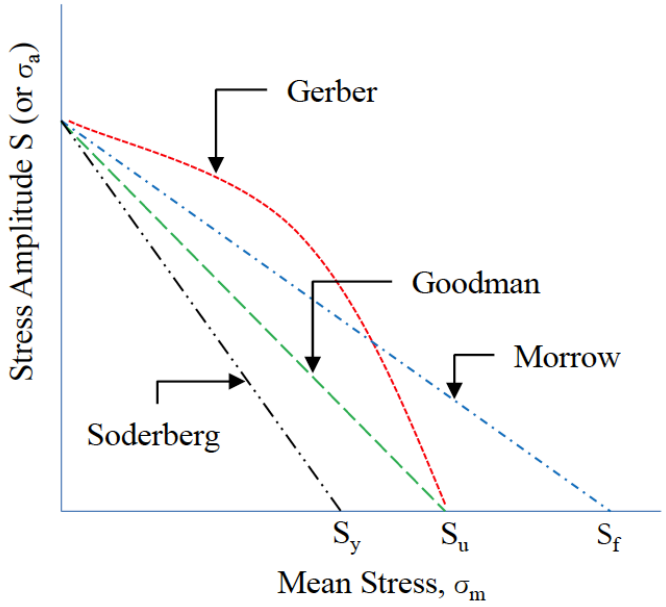


Figure 5.8.: Impact of mean stress on stress amplitude

of cycles:

$$W_i \propto n_i \times \sigma_{a_i} \tag{5.9}$$

And assuming the critical damage is the same across all stress levels:

$$W_{failure} \propto N_i \times \sigma_{a_i} \tag{5.10}$$

Therefore C, proportion of the cumulative damage to the critical value, can be calculated by:

$$C = \frac{\sum_{i=1}^{k} n_i \times \sigma_{a_i}}{W_{failure}} \tag{5.11}$$

When the damage fraction reaches 1, failure occurs.

There are several limitations to Miner's rule. It uses only the expected values and fails to recognize the probabilistic nature of fatigue. Furthermore, a simple linear life-stress relationship is assumed that may not be true in many real-word applications. To overcome the limitations of Miner's rule, there are other life stress models such as 'inverse power low model' that uses Weibull distribution in order to consider the probabilistic nature of the failures. Despite its limitations, Miner's rule can provide a simple but effective method for damage calculation.

In this study, a simplified implementation of this rule is used and impact

of mean pressure is considered in the development of the cumulative damage index (described in chapter 7).

Fatigue Corrosion and Stress Corrosion Cracking (SCC)

Corrosion fatigue is the mechanical degradation of the material under the combined action of corrosion and cyclic loadings. Various inter-granular/ trans-granular cracks can develop in low/high stress cycles of stress, impacted by various environmental factors, listed in Figure 5.9.

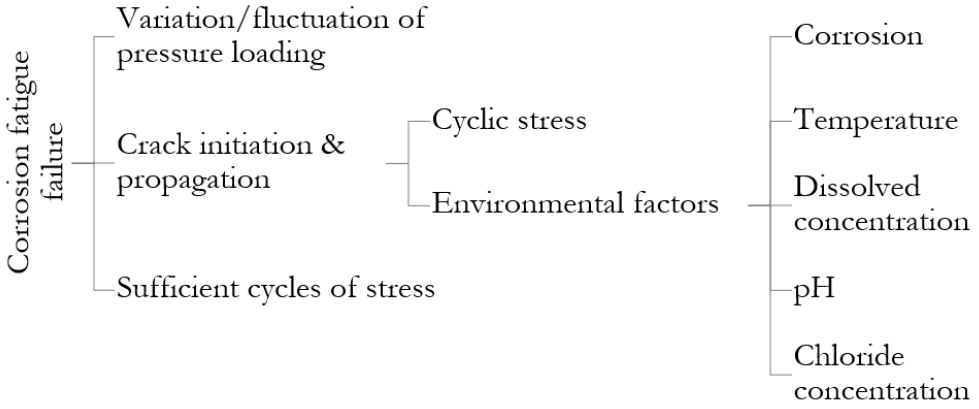


Figure 5.9.: Corrosion fatigue factors

In stress corrosion cracking, growth of crack formation is a result of the combined action of corrosion and tensile stress, resulting in various types of cracks, e.g. isolated or colony, inter-granular/ trans-granular cracks in high/low pH, etc. In this process, depth of crack and corrosion are combined causing reduction in pipe strength (see Section 2.3.2 for more details).

While in SCC, a specific combination of material, corrosive environment and tensile stress is required for the SCC to develop, for corrosion fatigue the only requirement is that the component is under tensile stress. Table 5.1 shows a comparison between characteristics of SCC and fatigue corrosion.

In a normal (non-corrosion) fatigue, a substantial length of the component life (stress loading cycles) is spent in crack nucleation. In comparison, crack propagation occurs in a much shorter period of time. In contrast, in corrosion fatigue, crack nucleation is facilitated by corrosion, and as such, only a fraction of the fatigue life is sufficient for this stage; majority of life is spent in crack propagation. Depending on the characteristics of the material and corrosive environment, this could result in, possibly a considerable, shift in $S-N$ curve. As shown in Figure 5.10, against the normal $S-N$ curve (curve A), this could result in elimination of fatigue endurance limit (curve B), or a whole shift of

Table 5.1.: Comparison of SCC and fatigue corrosion-adopted from Roberge (1999)

Factor	SCC	Fatigue corrosion
Stress	Static tensile	Cyclic + tensile
Aqueous corrosive	Specific to the alloy	Any
Temperature increase	Accelerates	Accelerates
Pure metal	Resistant	Susceptible
Crack morphology	Transgranular, Intergranular, Branched	Transgranular, Unbranched, Blunt tip
Corrosion products in cracks	Absent	Present
Crack surface appearance	Cleavagelike	Beach marks and/or striations
Cathodic polarization	Suppresses	Suppresses
Near maximum strength	Susceptible but minor	Accelerates

the curve lowering the fatigue strength (curve C).

Notches impact

The minimum stress amplitude to overcome the crack growth barrier for further crack propagation is referred to as *fatigue limit* (Lee et al., 2005). The fatigue limit can be negatively influenced by other factors such as periodic overload, elevated temperatures, irregularities in the material or corrosion.

Most fatigue failures, in practice, occur at notches or stress concentrations. The term notch might refer to holes, grooves, shoulders, keyways, etc. Such stress concentration factors often result in maximum local stresses $\sigma_{a_{\max}}$ that are many times greater than nominal stress σ_a of the component (Engrasp, 2008). For an ideally elastic member, K_T is defined as theoretical stress concentration factor:

$$K_T = \frac{\sigma_{a_{\max}}}{\sigma_a} \tag{5.12}$$

Fatigue notch factor (K_f) describes the effect of notches in the stress-life approach:

$$K_f = \frac{\sigma_{a_e}(unnotched)}{\sigma_{a_e}(notched)} \tag{5.13}$$

where σ_{a_e} is the fatigue strength (the endurance limit for ferrous metal). Unlike K_T, the fatigue notch factor, K_f, is dependant on the type of material and notch size. A notch sensitivity factor, q, explain the relationship between

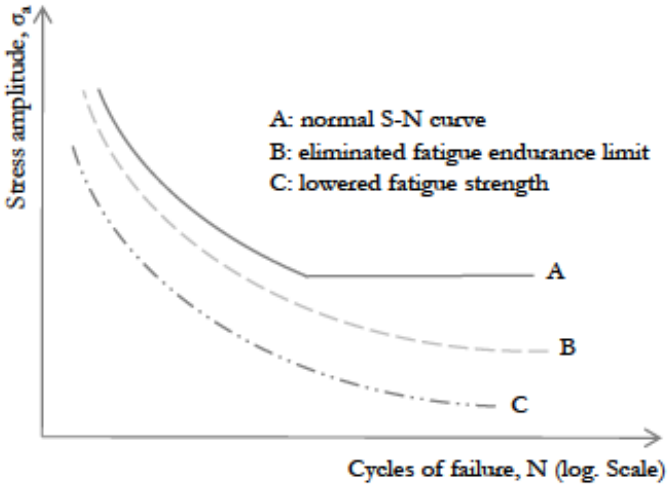


Figure 5.10.: Effect of corrosion on $S - N$ curve

these factors:

$$q = \frac{K_f - 1}{K_T - 1} \tag{5.14}$$

Based on correlation to experimental data, Peterson equation and Neubar equation are the most common ones developed in order to determine q. In Peterson equation (5.15), r is the notch root radius and a_N is material constant. Neubar equation (5.16) uses r, the notch root radius and ρ, a material constant related to grain size of the material:

$$q = \frac{1}{1 + \frac{a_N}{r}} \tag{5.15}$$

$$q = \frac{1}{1 + \sqrt{\frac{\rho}{r}}} \tag{5.16}$$

The notch factor K_f is usually used to correct the fatigue strength for the notched member:

$$\sigma_{a_e}(notched) = \frac{\sigma_{a_e}(unnotched)}{K_f} \tag{5.17}$$

Figure 5.11 shows an example for the impact of a 'notch' on a steel sample in the reduction of fatigue life, compared to a sample free from this.

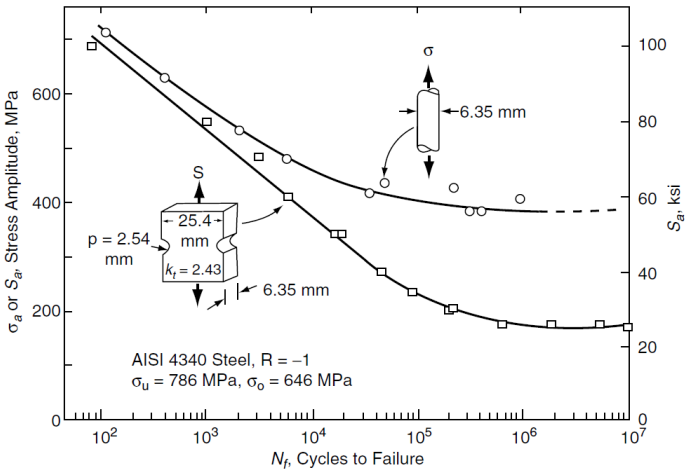


Figure 5.11.: An example for notch impact on fatigue life reduction-adopted from Davis (2004)

Fatigue in aqueous and corrosive environments

It is generally agreed that the presence of aggressive aqueous environment drastically reduces the fatigue life of metal (Bhat and Patibandl, 2011). Initiation or nucleation life of fatigue crack in a polished specimen may drop to as low as 10% of total fatigue life in a corrosive environment. Exposure to a corrosive environment, according to Bhat and Patibandl (2011), potentially has a similar effect on fatigue life as machining of sharp notch into the surface. A corrosive environment facilitates the development of geometric discontinuities on the surface. These discontinuities, or stress risers, then become the potential sites for origination of fatigue cracks upon cyclic loading.

As Bhat and Patibandl (2011) suggest, in a corrosive environment, several mechanisms have been proposed to describe crack initiation and propagation. For crack initiation, these include surface film rupture, pitting, strain effect and surface adsorption. Two main mechanisms assist in crack propagation, anodic dissolution and hydrogen embrittlement. It is important to note that no single mechanism can fully explain the crack initiation or crack propagation behaviour of metals in corrosive environment.

Crack closure

During the unloading part of a load cycle, the opened crack faces touch each other earlier at position A instead of B, as shown in Figure 5.12. On further reduction of stress from A, the touched surfaces start exerting compressive load, the effect of which is overcome by a part of the next cycle. In other words, part of the energy of a new cycle is utilized in overcoming the compressive effects

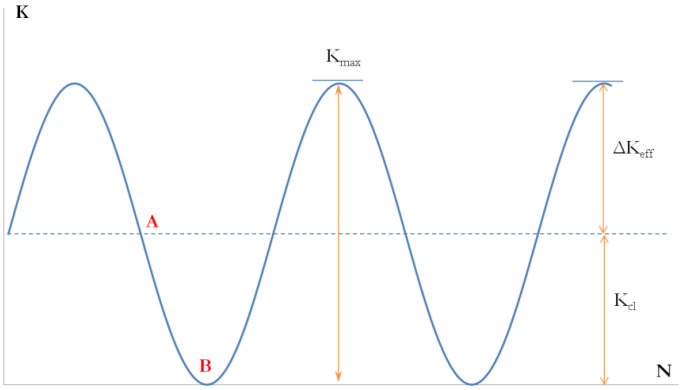


Figure 5.12.: Impact of stress cycles on crack closure

induced by the previous cycle. In the process, the applied cyclic stress intensity parameter, ΔK, reduces to effective stress intensity parameter, ΔK_{eff}, that is equal to $K_{\max} - K_{cl}$; and closure ratio defined as $U = \frac{\Delta K_{eff}}{\Delta K}$. Closure is induced by crack tip plasticity, surface oxides and roughness, etc.

Crack development in water pipes

In the case of water pipelines, the initiation phase of fatigue might be of little concern. A large number (range of 10^5 to 10^6) of hoop-stress cycles resulting from pressure fluctuations is required to cause failure. Such frequency is often far greater than the level most pipelines would experience.

It is, therefore, rare for pipes to experience fatigue failure in absence of initial flaws, damage or other fractures that concentrates stresses locally. However, existence of such flaws can dramatically change the component's behaviour. These flaws may have the effect of concentrating stresses by factors of 2 to 5 (Rosenfeld and Kiefner, 2006). Therefore, locally concentrated stresses may be great enough to cause failure in much fewer cycles, possibly thousands of cycles. In such cases, the initiation phase is bypassed and fatigue failure is governed by the fatigue crack propagation life.

The residual structural capacity of water mains is affected by corrosion, degradation and material deterioration due to environmental and operational conditions as well as quality of manufacturing and installation.

Cyclic loading could lead to sub-critical crack growth from corrosion-induced defects and eventually promote failure. Existence of cracks, as a result of these factors, substantially reduces the number of cycles required for development of fatigue failure.

Shown in the Figure 5.13, hoop stresses imposed by cyclic loadings of pressure fluctuation in the pipe continue to enlarge the cracks and expose a larger surface

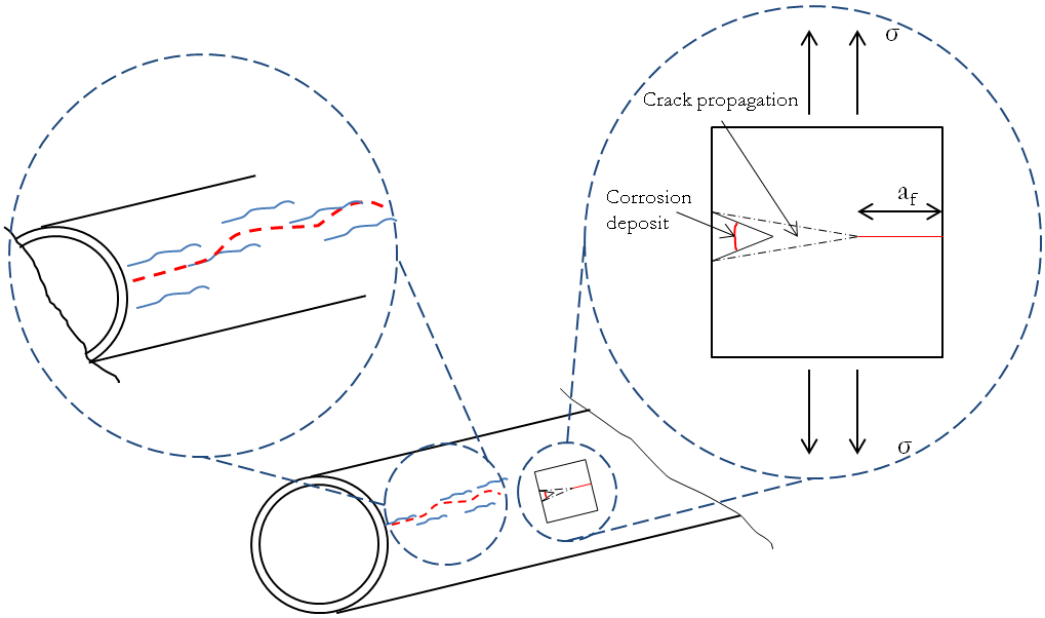


Figure 5.13.: Crack development caused by cyclic loadings on the pipe

of the material within the crack. Furthermore, when corrosion is developed in such opening, expansion of the corroded deposit enforces additional loadings on the inside of the crack, enlarging the opening even further. This in turn exposes more area to corrosive environment developing additional corrosion. This cyclic phenomenon combined with the crack development as a result of hoop stress will cause a speedier crack growth, in comparison with each factor impacting in isolation.

Crack development will continue until the residual wall thickness, shown as a_f in Figure 5.13, is low enough not to resist the next loading stress. If an individual crack is not sufficiently large to cause pipe failure, multiples of such cracks, when aligned, can join to develop larger cracks.

5.2. Cumulative Pressure-Induced Stress (CPIS)

Pipe failures for ductile material, as a result of internal pressure-related stresses, can occur in two ways. Firstly, failure can happen if the pipe is exposed to an extreme pressure transient event that exceeds the pipe's tensile strength. Such an extreme tensile stress can result in a fracture occurring in the plastic region of stress-strain curve of a mild steel, point B in Figure 5.14. If the pipe is exposed to a series of cyclic tensile loadings, individually not large enough to cause failure, the cumulative impact of such stresses can lead to fatigue failure of the pipe. Such failure may occur in the elastic region of the stress-strain

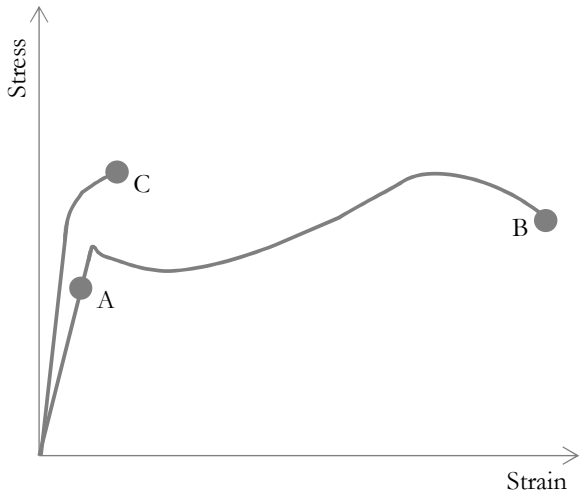


Figure 5.14.: Stress-strain curve for mild and brittle steel

curve, point A in Figure 5.14.

Low temperature (below 5°C) can increase embrittlement of cast iron pipes. Excess loading can then cause brittle fracture, point C in Figure 5.14.

5.2.1. Cumulative stress

The cumulative damage from any cyclic stress can be estimated by equation 5.11. Such cyclic stresses can include diurnal steady, quasi-steady and unsteady pressure variations. A metric, Cumulative Pressure-Induced Stress or CPIS, is proposed to represent the total stress imposed on the pipe from various cyclic pressure related loadings, as illustrated in equation 7.1, repeated here:

$$CPIS = f[P_{mean}, DP] \tag{5.18}$$

where P_{mean} is the diurnal mean pressure and DP is the quasi-steady state and unsteady state pressure variations (including pressure transient events), and:

$$P_{mean} = [mean(P)]_{24hours} \tag{5.19}$$

$$DP = \sum_{i=1}^{M} (\sigma_i \times n_i) \tag{5.20}$$

where σ_i is amplitude of pressure variations, n_i is the frequency of cycle σ_i events and M is the number of different cyclic loading events.

Fatigue damage is strongly associated with the cycle ratio, n_i /N_i, where

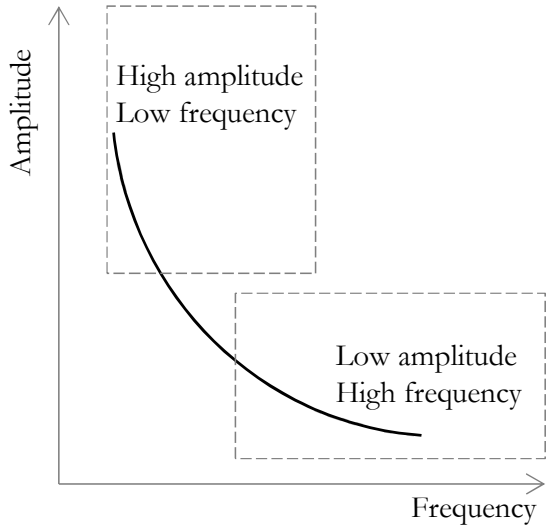


Figure 5.15.: Pipe failure due to pressure variations stress ranges

n_i is number of applied stress (or strain) cycles and N_i is the fatigue life at a combination of stress (or strain) amplitude and mean stress level. A water network is exposed to a large number of varied-sized pressure variations, both as diurnal pressure variations and pressure transient events of various amplitude and frequencies. The CPIS metric will represent the total stresses imposed from such loadings. A pipe failure can occur as result of this, when the system is either exposed to a large number of low amplitude stresses, a relatively small number of large stresses, or a combination of the two, Figure 5.15.

Limitations

The CPIS metric developed here has a number of limitations. Whilst it captures cumulative stresses from the combined impacts of multiple events with various characteristics, it fails to consider specific impacts of individual events with significant amplitude. Furthermore, CPIS does not provide information about the distribution of events over the logging period, i.e. events occurring over a short period of time with no events at other times, scattered events over the logging period or a fairly uniform distribution of events. In general, these are important for estimating cumulative damage, however, capturing additional information could provide insight into potential sources of hydraulic instabilities. In addition, regular and irregular cyclic events could have impact on behaviour of crack propagation and crack closure.

Furthermore, an individual pipe's condition (specifically level of corrosion

and potential cracks) are not considered in CPIS. Stress concentration regions and existence of cracks could significantly accelerate pipe's degradation and fatigue failure. However, lack of high quality data on mains condition across the network is a significant barrier preventing inclusion of this factor effectively.

A recent study (Stoianov and Sadowski, 2015), analysed fatigue failure of CI samples from a recently failed water pipes for various rate of loading. The rates of loading imposed on water pipes from pressure transient events were found to be significantly smaller than the rates that can alter fatigue behaviour of CI material. Therefore, rate of loading is not considered as a factor in the analysis in this study.

Depending on the load sequences (low to high or vice versa), there could be increase or decrease in the cumulative fatigue damage. Moreover, CI materials used in water pipes range significantly in quality, performance and structural characteristics. Therefore, further investigation in this area is needed to evaluate the impact from changes in loading characteristics.

5.3. Summary

In this chapter, the fundamentals of fatigue failure, its primary characteristics, fatigue mechanisms and analyses as well as factors affecting fatigue were reviewed. CPIS, a cumulative pressure-induced stress metric, was introduced, representing the network's hydraulic instabilities causing material degradation and ultimate fatigue failure.

In order to develop CPIS for a network's pipes, high frequency pressure data is needed across the network. In the following chapter, a systematic method is presented in order to drive the high frequency pressure data collection experiment.

6. Experimental programme

Studying the association between pipe failure records and the network's hydraulic characteristics relies upon an experimental sampling programme for the acquisition of data with unique spatial and temporal resolution.

In this chapter, design, development and execution of the research experimental programme is provided. This includes a sampling survey developed to collect high frequency pressure data across the network as well as a pipe burst data collection programme, designed to gather data on burst events and repairs activities.

The objective of this chapter is to explore and develop the selection and design of the sampling survey methodology, including the criteria and analysis required to determine sample sizes, area categorisations and ranking, random sampling and experiment's execution and verification.

The chapter is structured as following. Firstly, a background review is provided on appropriate data collection methods, followed by a review of various categories of probability sampling methods. Sample sizes and sampling errors and biases are then discussed and calculation of sample sizes for this sampling survey is presented. Stratified random sampling is reviewed with development of a two-stage sampling survey procedure. The following sections discuss the pressure logging duration and execution of sampling exercise. This section finishes by providing a verification of network representations achieved by recorded samples.

6.1. High resolution pressure data sampling

Long term impact of pressure variations on water pipes are not well understood. One of the main reason for this could be attributed to the shortage of high frequency pressure data. Lack of such data is mainly due to the lack of adequate instrumentation to gather operational high frequency pressure data and analytical tools to analyse collected data.

Recent technological developments have made it possible to capture and analyse high frequency pressure data across the system effortlessly. A platform developed at Imperial College London, *InfraSense* from Hoskins (2015), is

used for monitoring the hydraulic behaviour across various assets of a water supply networks. High frequency pressure logging (up to 128 samples per second) is undertaken at **480** monitoring locations across a water supply network, by the researcher and support of the Company's technicians. A data collection programme is designed to capture assets' characteristics and hydraulic performance. Design, development and execution of this programme is detailed below.

6.1.1. Data collection and sampling

Data collection methods include census, sample survey, experiment and observation study. *Sampling survey*, unlike *experiment*, does not require control over the subject group. It is also suitable for generalisations, unlike *observation* study. Amongst these, sampling survey is the most suitable for collection of pressure variation data. To be able to generalise the finding to a larger population, *random* selection is required.

There are two categories of sampling methods: probability and non-probability sampling methods. In probability samples, each element from the population has a known and non-zero chance of being chosen. In non-probability samples, it is not known for certain whether each element has non-zero probability of being selected. Some of the most common categories of probability sampling methods include the following (Fuller, 2011):

- Simple random sampling
- Stratified sampling
- Multi-stage sampling
- Cluster sampling
- Systematic sampling

In *simple* random sampling, desired number of subjects (sample size) are randomly selected from a list that includes all members of the population. *Stratified* sampling is where the subjects are grouped in a number of classifications, and then a final list of subjects is randomly selected from different strata. In *systematic* sampling, every *i-th* member of a list that includes all members of the population are selected. The first step in *cluster* sampling involves identifying a number of clusters within the population. Then, a few identified clusters are selected. At the next step, depending on the study goals,

all or a subset of randomly selected members of each cluster are selected. *Multi-stage* sampling involves using a combination of two or more of the sampling techniques highlighted above.

6.1.2. Stratified random sampling

In stratified sampling, the population is divided to a number of groups (strata), based on certain characteristics. Then, within each group, a probability sample is undertaken to select the samples. Compared to cluster sampling, though both methods use a non-overlapping subset of population, in stratified sample, all strata are presented in the final sample while only a subset of clusters are included in the final sample. In most cases, stratified random sampling is the preferred method in comparison to simple random sampling and, when elements in strata are internally homogeneous, in comparison to cluster sampling. Compared to simple random sampling, stratified random sampling provides a greater precision, requires smaller sample size and guards against under-representative samples (e.g. all elements of a population with similar characteristics could be missed in a simple random sampling). Stratified sampling is selected as the suitable sampling method for this study.

Two categories of stratified random sampling are: proportionate stratification and disproportionate stratification (Hansen et al., 1993). In proportionate stratification, sample size of each stratum is proportionate to the population size of it, i.e. each stratum has similar sampling fraction (ration of sample size to population size). Sampling fraction may vary from one stratum to the next in disproportionate stratification. Disproportionate stratification usually provides a more suitable choice (e.g. lower cost and high precision) if sample elements are assigned correctly to strata. Various assumptions need to be made and the analysis' goals and their priorities to be considered in order to take advantage of this method. For instance, using this method, there could be sampling more heavily from stratum where stratum's population size is large or variability in stratum is large, etc.

6.1.3. Survey sampling bias and errors

A sample is considered a good sample if it is representative. *Bias* is the tendency of a sample statistics to systematically over or under estimate a population parameter. The bias that results from an unrepresentative sample is a *selection bias*. Examples of selection bias include *under-coverage* (where some population members are inadequately represented), *non-response bias* (e.g. where some locations cannot be logged), and *voluntary response bias*

(sample subjects are self-selected-this is not applicable in this study). Random sampling is based on chance and a known non-zero probability of each member being selected. This protects the sampling against under-coverage bias. Other types of bias are due to measurement errors (resulting in response bias) and from poor measurement process (e.g. survey environment, etc.)

A sample is a selection of cases chosen in a way to represent the total population as well as possible. The quality of sample survey results are measured in i) accuracy, ii) precision, and iii) margin of error. *Accuracy* is an indication to how close a sample statistics is to its population parameter. *Precision* measures how close are estimates from difference samples to one another. Precision of a sample design is directly related to variability of estimates; two common measures of variability are standard deviation and standard error of estimate. *Margin of error* expresses the maximum expected difference between true population parameter and a sample estimate of that parameter. Confidence interval is the probability that X% of the entire population actually lies within borders of margin of error (Box et al., 2005).

6.1.4. Sample size

There are various factors affecting the size of a sample including: minimum acceptable level of precision, confidence interval, variability within population or sub population (e.g. clusters, strata, etc.), cost consideration, administrative concern (complexity of design, research time-frame, etc.), sampling method, etc.

Considering these factors, in order to calculate the sample size for random sampling without replacement, population size, confidence level, margin of error and standard deviation are required, equation 6.1. Random sampling without replacement usually provide equal or better precision in comparison to the sampling with replacement. For a large population, sample size can be calculated as (Hansen et al., 1993):

$$n = (z - score)^2 \times SD \times (1 - SD)/(MoE)^2 \tag{6.1}$$

where n is sample size, SD is standard deviation and MoE is margin of error. Value of 0.5 can be assumed for standard deviation when unknown. This is a conservative estimate resulting in sample size producing at least the required precision.

For stratified random sampling, the sample size in each stratum can be calculated using equation 6.2 and 6.3, for proportionate and disproportionate

methods, respectively (Thompson, 2012).

$$n_h = (\frac{N_h}{N})n \tag{6.2}$$

$$n_h = \frac{n \times \frac{(N_h \sigma_h)}{\sqrt{c_h}}}{\sum_{i=1}^{N_s} \frac{(N_i \sigma_i)}{\sqrt{c_i}}} \tag{6.3}$$

where n_h is the sample size for stratum h, n is total sample size, N_h is the population size for stratum h, N is the total population size, N_s is the number of strata, σ_h is the standard deviation of stratum h, and c_h is the direct cost to sample an individual element from stratum h. In this study, it is assumed that direct cost of each sample is the same across various strata, and in order to maximise precision given a fixed sample size, *Neyman allocation* (Fuller, 2011), can be used to provide the sample size for stratum h as:

$$n_h = \frac{n \times (N_h \sigma_h)}{\sum_{i=1}^{N_s} (N_i \sigma_i)} \tag{6.4}$$

6.1.5. Sample size calculation

There are approximately 24000km length of metallic pipe in the study areas, associated to approximately 0.5million pipe segments, with a large proportion $< 5m$ in length. For the purpose of total population size, it is assumed that a logging location can be representative of the hydraulic characteristics for a pipe length of 1km, 500m length upstream and downstream direction. This will lead to a total population of 24,000.

In order to calculate the sample size, $z - score$ and margin of error are assumed based on the requirements of the research's goals. Using equation 6.1, and considering a standard deviation of 0.5, sample sizes (n) are calculated for various confidence intervals and margins of error, as presented in Table 6.1.

Table 6.1.: Sample size for various confidence intervals and margins of error

α	Confidence interval	z-score	Margin of error	SD	n
0.05	95%	1.96	10%	0.5	96
0.02	98	2.33	10%	0.5	136
0.1	90%	1.645	5%	0.5	271
0.05	95%	1.96	5%	0.5	384
0.02	98%	2.33	5%	0.5	543

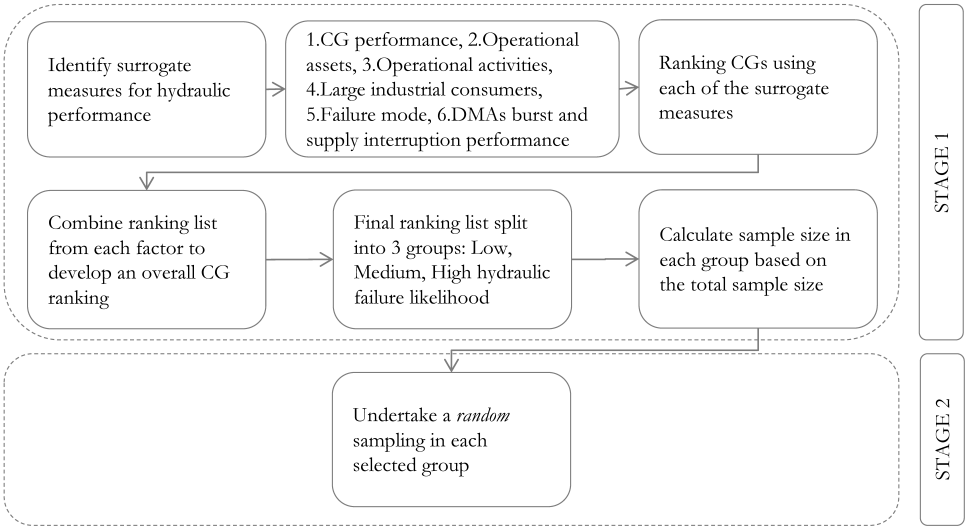


Figure 6.1.: Stages of stratified random sampling

Considering the network size, uncertainties in data used for sampling design, and limitations on time, equipment (20 loggers) and logger deployment resources available to this research work, confidence interval of 95% and margin of error of 5% are considered, resulting in a sample size of 384. It should be noted that this is the minimum sample size. In practice, in order to provide coverage of the areas identified for logging, additional logging might be undertaken, as detailed in the following section.

6.2. Stratified random sampling

The stratified random sampling consists of two stages. Stage 1 includes assigning the population members to individual stratum, and in stage 2, random sampling at each stratum is undertaken, based on the sampling size in each stratum, Figure 6.1.

In the following sections, the process to develop three strata is explained, followed by a random sample selection in each stratum.

6.2.1. Stratified random sampling - stage 1

In stage 1, three strata with *low*, *medium* and *high* hydraulic performance index are defined. This is to ensure network coverage for different characteristics, taking into account the importance of network's hydraulic and performance indicators. *System hydraulic performance index* is reviewed by analysing data on asset and historical performance measures and employing spatial statistics

and distribution, cluster analysis, hot spot analysis and value ranking.

A study of pressure transient impact should be undertaken within the network's subsystems that contain hydraulically connected assets. Such areas should be larger than DMA area, as sources of majority of pressure variations and transient events are outside DMAs. For the purpose of developing the strata, the analyses are mainly undertaken at the CG level. Some hydraulic measures such as pump switch or DMA performance are done at pump or DMA level. In such cases, associated CG is used for the categorisation, as described below.

System hydraulic difficulty is defined to include hydraulic-related factors that result in operation complexity and/or partial delivery failures. These are historical pipe failure, leakage level, supply interruption incidents, possibility of pressure transient occurrence, etc. Areas and assets with high probability of pressure transient occurrence include operational sites/areas historically exposed to pressure transient events (based on previous studies, local knowledge, operational expertise, etc.), operational assets (e.g. pumps, valves, etc.), large industrial consumers, PRVs, areas with large volume of operational activities, etc. Factors used for stage 1 ranking are:

- Control Group (CG) ranking
- Operational assets
- Operational activities
- Large industrial consumers
- Failure mode
- DMAs burst and supply interruption ranking

In addition to these, other system operational difficulties (such as leakage) were originally considered, however, preliminary analyses indicated those were not suitable indicators. Background leakage could be significantly impacted by the fluid dynamics. It can also have considerable impact on the energy dissipation of pressure variation and transient waves across the network. However, in current ranking exercise, leakage levels were not considered as reliable indicators of DMAs performance and pipe failure rate. Firstly, considerable discrepancies between DMAs' leakage indicators from various data sources as well as significant gaps in available data result in an incomplete datasets for leakage performance comparisons in DMAs. Furthermore, leakage levels are

often affected by numerous other factors, such as illegal use, meter errors, etc. and as such not a reliable indicator of actual pipe failures within the DMA.

The analytical process for each factor is described below.

Control Group ranking

Based on normalised burst rate per length of mains in each CG (number of historical bursts/1000km length of main), all 130 CGs in STW area are ranked. This provided a league table of the highest to least offending CGs, based on historical mains failure (Table 6.2). CG levels in this table are developed by equally dividing all CGs into three levels (high, medium and low).

Table 6.2.: Example of CG ranking based on CG pipe failure rate

CG reference	Number of burst (2003-2013)	Length of mains(km)	Bursts/ 1000km	CG ranking	CG level
CGEM23	837	246.9	3390	3	H
CGEM16	246	75.5	3258	4	H
CGEM03	288	152.7	1886	51	M
CGEM24	2840	1521.8	1866	52	M
CGEM47	434	321.9	1348	96	L
CGSV03	242	180.7	1339	97	L
...	...	...	...	...	...

Operational assets

Operational assets, specifically pumps' switches are amongst the most important factors in generating hydraulic instabilities across the network. Therefore, areas with high number of operational assets are likely to experience higher level of hydraulic instabilities. Using the SCADA system and available pump data, all pumps and boosters across STW are reviewed, and an analysis is undertaken to provide a league table based on the number of switches (on/off status changes) over a period of one year (2012).

Using this ranking and other information about each pump station (size, flow and number of pumps), control groups are ranked based on number of annual assets' switches, as shown in Table 6.3.

Operational activities

Operational activities on the network (e.g. valve operation) can generate hydraulic instabilities in the network. During the period of June 2012-June 2013, approximately 26,890 operational activities are recorded across STW

Table 6.3.: Example of CG ranking based on their operation sites

Site reference	Number of start/stop	Number of hours run	Switch per hour run	Pump rank	Site's CG	CG level
OS1218	1,826	51	17.9	18	CGSV03	H
OS1224	572	31	9.23	24	CGST14	H
OS1240	1,018	107	4.75	40	CGEM19	M
OS1270	220	162	0.68	70	CGSV37	L
...	...	...	...	...	...	...

Table 6.4.: Example of CG ranking from their operational activities

DMA reference	Number of operational activities	CG reference	CG level
1628	50	CGSV15	H
4247	48	CGEL04	H
6228	27	CGSV42	M
8121	20	CGSV12	M
12302	17	CGEM01	L
14429	16	CGEM15	L
...	...	...	...

network. Activity types include mains repairs, mains new installation and renewal, flushing, asset inspection and servicing, etc. These records are reviewed for quality checks and then spatially analysed to find their associated DMAs. CGs containing these DMAs are then ranked, as illustrated in Table 6.4.

Large industrial consumers

Some large industrial users extract large volume of water over a relatively short period of time. These often lead to changes in the flow/ velocity regimes resulting in pressure variations. It is therefore important to review their performance and potential impacts of such operation on the networks' assets.

Despite consuming large volume of water, some large consumers do not follow such demand profiles. For instance, health centres, hospitals and university residential halls are amongst the significant consumers, but due to their consumption pattern, risk of significant pressure variations generated by their consumption is relatively low and should not be included in the below analyses.

Analysing the billing data, a list of all large consumers across the Company is extracted and analysed using GIS's spatial analysis tools. CGs are ranked based on the number of large industry consumers in each CG, as per Table 6.5.

Table 6.5.: Example of CG ranking based on their large industry consumers

CG reference	Number of large industry consumers	CG level
CGEL02	12	H
CGEM24	14	H
CGEM18	2	M
CGSV34	3	M
CGEM01	≤1	L
CGEM02	≤1	L
...	...	...

Table 6.6.: Example of CG ranking based on type of mains burst

CG reference	Longitudinal bursts	Mains length(km)	Longitudinal burst/1000km	Longitudinal burst ranking	CG level
CGEM31	17	85.5	199	7	H
CGEM03	26	152.7	170	12	H
CGEM15	13	178.3	73	61	M
CGST05	16	249.8	64	77	M
CGEM45	36	898	40	111	L
CGSV20	9	300.8	30	123	L
...	...	...	...	...	...

Failure modes

Information about failure modes are often not recorded, when broken mains are repaired. In most cases, however, the available data about the mains and repair activities can be used to help determining the failure mode. Information about the mains material, size and repair method is used to group the mains failures to longitudinal and circumferential failures. For instance, replacing a length of a metallic pipe (e.g. instead of a collar repair) suggests that the failure has been a longitudinal mode. Different failure modes suggest the failure mechanisms and underlying factors and loading that resulted in the failure, as discussed in chapter 2. Pressure-related hoop stresses exposed on the pipe can contribute to longitudinal failures. Table 6.6 illustrates ranking of control groups based on density of longitudinal failures within each control group.

DMA burst and supply interruption ranking

A total of approximately 3200 DMAs are ranked with respect to supply interruption and historical mains burst. The ranking is developed based on the historical bursts records from year 2003-2013, supply interruption data for year

2012, DMAs property impact (number of properties at risk of supply interruption), DMAs KPI impact (a measure of the level each DMA can impact supply interruption's Key Performance Indicators), and DMAs annual interruption frequency. DMAs (and subsequently their associated CGs) are then ranked based on their performance, as illustrated in Table 6.7.

Combining control group rankings

In order to achieve a final control group rank list, six control group ranking lists, as developed by using the indicators explained in section 6.2.1, are merged. Mainly developed in the *voting theory*, there are a number of methods (e.g. Condorcet method, Schultz method, Voting System, etc.) that can be used to merge various ranked lists. Merging methods without learning mapping functions include Round Robin, Raw Score and Linear Scaling (Lin and Hauptmann, 2004). In *Round Robin*, top-ranked elements from all ranked lists are selected and included in the merged list, followed by selection of second-ranked elements, and so on. In *Raw Score*, local score for each list (rather than their ranking) is used for defining the ranking in the merged list. *Linear Scaling* is used to normalise the local score into the range between zero and one, used to populate the merged table.

For this research, a Raw Score method is used, where control groups are assigned a weighting based on their ranking in each list (CG ranking, DMA ranking, etc.). Then, a linear combination (total) of weightings for each control group is calculated and used to develop a final ranking list (Table 6.8).

Stratum sample size

Three strata of the CGs with *low*, *medium* and *high* level of hydraulic system difficulties developed in stage 1 are summarised in Table 6.9. Stratum with low hydraulic system difficulties represent areas of the system that have not experienced system hydraulic difficulties (e.g. pipe failure, supply interruption, etc.) and/or where occurrence of pressure variations and transient events are deemed to be less likely. Therefore a lower level of sampling is taken at these areas. Similarly, a larger proportion of sample size are assigned to the stratum with high level of hydraulic system difficulty.

Equation 6.4 is expanded to account for this, including a *sample adjustment weighting*:

$$n_h = \frac{n \times (N_h \sigma_h) \times SW}{\sum_{i=1}^{N_s} (N_i \sigma_i) \times SW_i} \tag{6.5}$$

Table 6.7.: Example of cg ranking based on the performance of their DMAs

DMA	CG reference	Number of bursts (2003-2013)	Mains length (m)	Bursts/ 1000km	DMAs Property Impact	DMA KPI Impact
14313	CGEM14	34	8151	4171	2861	1
6345	CGSV43	157	45648	3439	99	97
2526	CGST17	283	47410	5969	269	0
6910	CGSV38	23	19344	1189	636	0
2856	CGSV23	55	20048	2743	196	142
1450	CGOE03	116	74104	1565	646	6
...	...	...	...	...	...	...

DMA	Number of times DMA interrupted	DMA ranking	CG ranking	CG level
14313	6	224	12	H
6345	4	388	51	H
2526	5	68	68	M
6910	4	2057	72	M
2856	5	705	117	L
1450	6	1653	127	L
...	...	...	...	...

Table 6.8.: Combined rankings generating overall CG ranking (W=weighting; L=level)

CG reference	CG ranking		DMA ranking		Operational activities		Pump switch		Burst type		Large consumer	
	W	L	W	L	W	L	W	L	W	L	W	L
CGEL04	23	H	19	H	7	H	43	L	99	L	5	H
CGEM11	46	M	6	H	5	H	51	L	49	M	7	H
CGST16	113	L	103	L	46	M	10	H	83	M	40	M
CGSV04	9	H	104	L	58	M	28	M	122	L	106	L
CGSV25	80	M	123	L	108	L	112	L	29	H	118	L
CGSV26	74	M	78	M	121	L	113	L	129	L	119	L
...	...	...	...	...	...	...	...	...	...	...	...	...

CG reference	Combined weighting	Overall CG Ranking	Overall CG Level
CGEL04	196	3	H
CGEM11	164	1	H
CGST16	395	73	M
CGSV04	427	82	M
CGSV25	570	119	L
CGSV26	634	126	L
...	...	...	...

Table 6.9.: Characteristics of the Strata resulted from stage 1

Stratum	Number of control groups	Population (pipe segments)
Low	43	477,304
Medium	43	318,018
High	43	456,905

Table 6.10.: Distribution of sampling points in each stratum

Stratum	Mains length(km)	Population (pipe segments)	Sample adjustment weighting	N_h	n_h
Low	10,193	477,304	0.2	95461	62
Medium	13,909	318,018	0.3	95405	62
High	19,473	456,905	0.5	228453	148

where n_h is the sample size for stratum h, n is total sample size, N_h is the population size for stratum h, σ_h is the standard deviation of stratum h and SW is sample adjustment weighting. As standard deviation for each stratum is not known, value of 0.5 is used. For n=384, σ_h =0.5 and considering population proportion (weighting factor to promote areas with higher risk); number of samples in each stratum is calculated, as shown in Table 6.10.

6.2.2. Stratified random sampling - stage 2

At stage 2, a *random* sampling is undertaken for the number of samples in each stratum. Logging points are randomly selected across the system in each stratum, taking into account various practical constraints and deployment restrictions. Some of these restrictions include:

- lack of access points on large diameter (trunk) mains
- lack of operational and log-able tapping points upstream/downstream of valves and pumps
- restriction of access to certain locations (traffic management, private properties, etc.)
- discrepancies between available asset data and the real network configurations

6.3. Logging duration

Traditionally, due to technological limitations (mainly battery life and memory size), high frequency loggings were undertaken only for a limited period of time (i.e. a few hours and up to a few days). Utilising the latest development in data collection and handling, using the InfraSense platform, each point is logged for a minimum of two weeks.

For steady-state hydraulic modelling, Savic and Banyard (2011) recommends a logging duration of one full week (and longer if necessary) to be sufficient to account for different demand and flow profile.

Due to various logistical and practical limitations, any logging activity, at best, can provide a snapshot of the network's properties during the logging period. This is a limiting factor specifically regarding the network's hydraulic characteristics, as this can change depending on the time of the year. In order to investigate this limitation, in 10 selected locations, the same assets were logged both in winter and summer periods, to understand potential changes in dynamic of the network at different demand and operational regime.

Furthermore, networks hydraulic characteristics depends on the operational regimes, network configurations, consumption patterns (e.g. due to different pattern during holidays), etc. Longer period of loggings at various locations are undertaken to assess annual variations. Approximately 60 locations were logged for around four weeks and in 30 locations for a period of 20 weeks. In most cases, pressure profile and fluctuations follow similar pattern during the logging period. Analysing the dataset collected at these locations confirmed that a two-week logging period could provide a good representation of the entire year. Depending on the network characteristics, there might be locations where pressure profile is highly season-specific (e.g. areas with hotels that have particularly high demand during peak seasons). In such cases, additional logging is required to confirm the varying profile.

In the studies by Collins (2015), several locations were logged using high frequency pressure loggers for 11 months intermittently. This was to confirm variability of pressure profile, specifically quasi-steady and unsteady pressure changes. With regards to minimum required duration of logging, their findings are aligned with the findings of this research work.

Review of the system changes logs and discussions with the network operational teams confirm that, over the last 20 years, most hydraulic areas remain relatively unchanged. Most changes have been mainly limited to DMAs boundary changes, reconfigurations, rezoning and source alteration for a limited DMAs. Furthermore, many of such changes are also temporary and the

network operation is restored following resolution of the needs driving the change. Therefore, it can be safely assumed that amplitude and frequency of hydraulically-originated stresses remain relatively constant. Furthermore, over the last two decades, most supply systems have seen hydraulic improvements due to significant increase in the level of pressure control (reduction) activities. Consequently, water supply networks were subject to more extreme maximum pressure and hydraulic instabilities in the past. Therefore, the systems were exposed to more damaging pressure fluctuations in the past and data collected in this study is on the conservative side.

Despite the above, the relatively short time-span of the recorded data and assumptions on the pressure profile remaining unchanged are important limitations and should be considered when interpreting the results. Further long term and permanent logging installations are required in order to continuously capture, characterise and quantify the fluid dynamics for the purpose of providing further evidence on the hydraulic changes across the network.

6.4. Execution of the experiment

InfraSense, developed at Imperial College London, provided a sophisticated platform for capturing and analysing high frequency pressure data (details in chapter 3). The platform provided a reliable deployment procedure with very limited deployment failures and a robust data management and visualisation capabilities.

Extensive planning and management was made in order to overcome various logistical challenges when deploying pressure loggers. Specifically, site visits and deployment of loggers at operational sites and customer properties required detailed planning and work procedures, in order to secure access to these sites and undertake the work in a safe manner complying with various legislations and guidelines. Furthermore, extensive communication was made with relevant operational teams when planning and executing the logger deployment and also for interpreting the collected data. Further details about the deployment processes is described in section 3.4 in chapter 3.

There were a number of challenges for undertaking the experimental programme, mainly due to limitations on the required support for logger deployment and also due to weather-related restrictions, access issues, locating underground assets, etc. In some cases, interventions were required in order to find alternative suitable locations for logging.

During logger deployment and collection, additional information about the assets as well as hydraulic and environmental conditions of the targeted assets

were collected. These include type and model of the assets (e.g. PRVs), assets operational regime (e.g. pumps), pressure reading at deployment and collection time, assets elevation, etc. recorded in the form shown in Figure A.1 in the appendix A.1. In order to increase data collection efficiency and accuracy, a tablet application was developed to collect these data as well as gathering geo-referenced photographic survey of logging locations and associated assets (e.g. pump, PRV, etc.). Such data was then used to investigate the network's operational characteristics, identifying sources of pressure transient events, impacts of various assets on pressure fluctuations, etc.

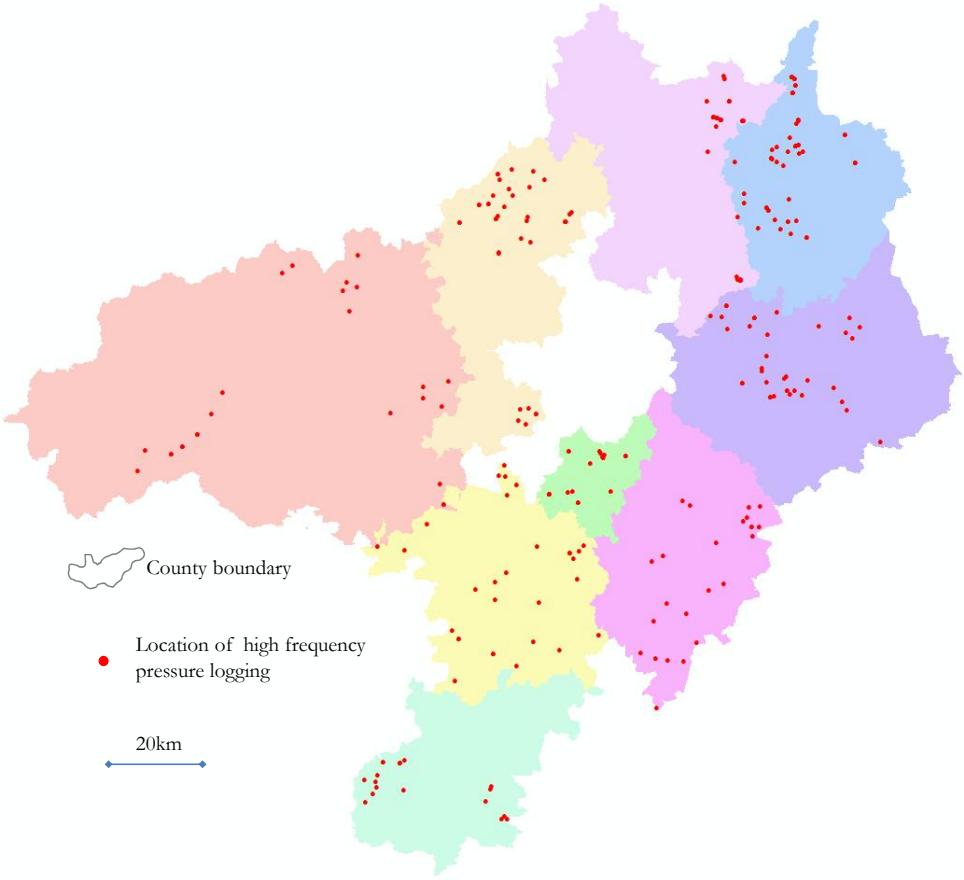
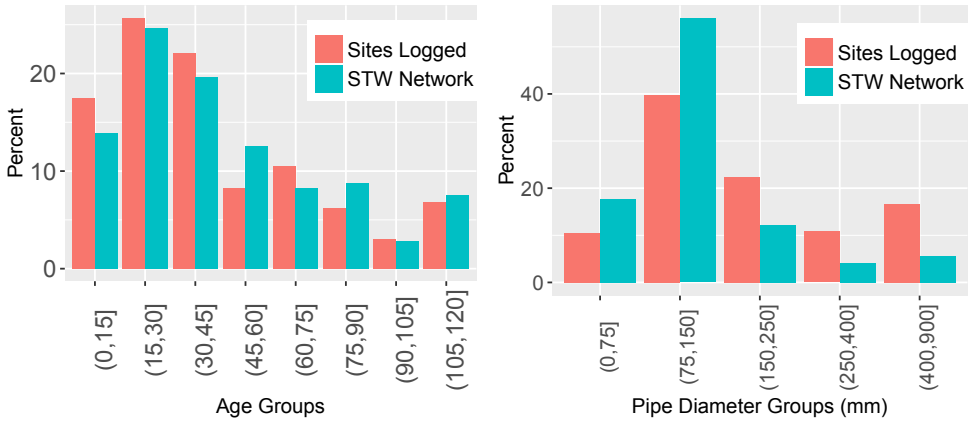


Figure 6.2.: Locations of high frequency logging mapped on counties within the Company's region

6.5. Verification of network representation

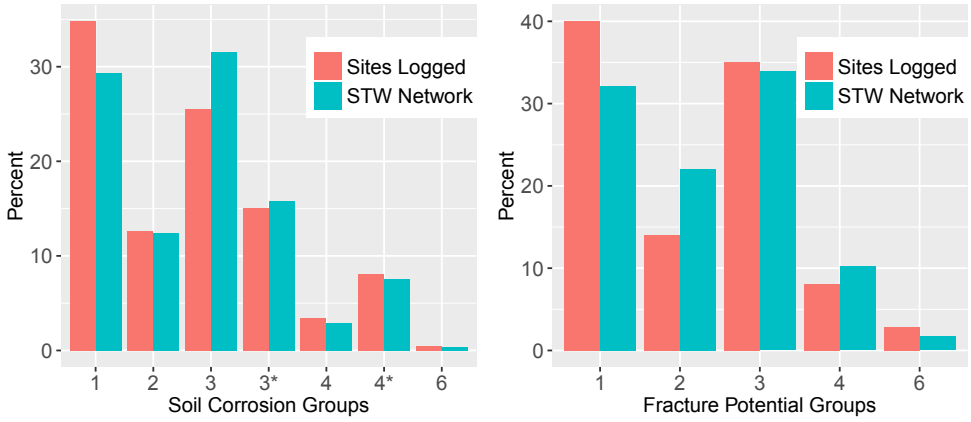
Exceeding the minimum sample requirement of 384, a total of 480 sites were logged during the sampling programme. Figure 6.2 shows the distribution of logged sites across various areas of the Company.

A number of reviews and analyses are undertaken in order to investigate the network's representativeness of the logged locations. Analyses are carried out to examine the distributions of logged assets. This is to examine whether these distributions follow similar pattern as distributions of all assets for the entire network). As illustrated in Figures 6.3, 6.4, 6.5 and 6.6, the distribution of sites logged, often closely, follows the same distribution of the entire network's assets. This is true for both physical properties (e.g. pipe age and soil properties) of the network and its hydraulic characteristics (e.g. maximum pressure).



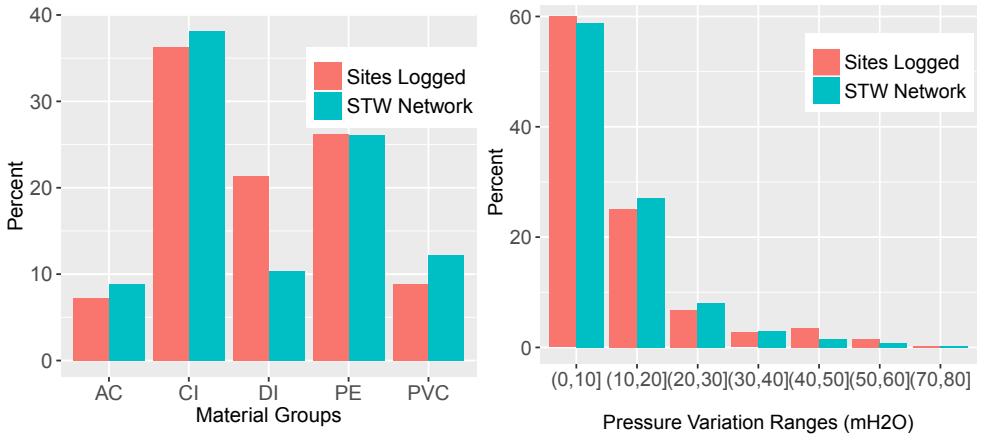
(a) (b)

Figure 6.3.: Comparison of distribution for assets logged versus all pipes (entire Company) at (a) age groups (b) pipe diameter groups



(a) (b)

Figure 6.4.: Comparison of distribution for assets logged versus all pipes (entire Company) at (a) soil corrosion groups (b) fracture potential groups



(a) (b)

Figure 6.5.: Comparison of distribution for assets logged versus all pipes (entire Company) at (a) material groups (b) pressure variation ranges

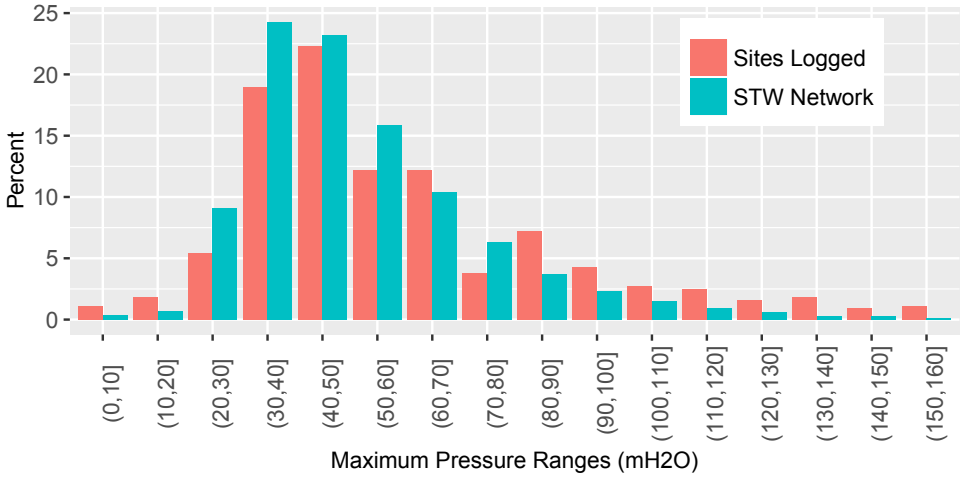


Figure 6.6.: Comparison of distribution for assets logged versus pipes (entire Company) pipes at maximum pressure ranges

6.6. Summary

This chapter summarised design, development and execution of a sampling programme to collect high frequency pressure data across the network. Stratified random sampling with a two-stage sampling survey procedure is used to design the sampling locations. In the next chapter, analysis of the collected data is discussed.

7. Pressure data analysis and CPIS development

A total of 480 locations are logged to gather high frequency pressure data, as described in chapter 6. Collected data are analysed in order to estimate the amplitude and frequency of the pressure cycles that collectively contribute to pipes deterioration. These metrics are then estimated across various nodes inside the network, based on the collected data at upstream nodes, and considering various forms of energy dissipation mechanisms (described in the following sections).

In this chapter, firstly, by reviewing the collected data, various categories of hydraulic instabilities and pressure profile distributions are described, based on their amplitude and frequency, with examples given from each group. Then, extraction of individual cyclic events from the high frequency pressure data is discussed with details about the cycle counting algorithm used. Following section describes the hydraulic tracing algorithm and methodologies used to calculate the shortest hydraulic path and characteristics of the pipes between two nodes in the network. Finally, methods are introduced to estimate energy dissipation of pressure fluctuation waves in order to estimate Dynamic Pressure, DP (described in section 5.2), at target nodes in the system.

7.1. Categories of pressure profiles

With regards to hydraulic profile and hydraulic instabilities, review of the collected data in this study suggests that the pressure distributions at network nodes often follow characteristics of one of these six categories:

- Group 1: Steady-state pressure profile with minimum instabilities ($<5mH_2O$ pressure change)
- Group 2: Predominantly steady-state pressure profile, with moderate (5-15mH_2O) pressure changes
- Group 3: Predominantly varying pressure profiles with frequent high amplitude (15-25mH_2O) pressure changes

- Group 4: Steady-state pressure profiles with infrequent but large ($>25mH_2O$) isolated transient events
- Group 5: Steady-state pressure profiles with frequent and large ($>25mH_2O$) transient events
- Group 6: Mixture of varying pressure profile with small to large transient events

Examples from these groups are presented in Figures 7.1, 7.2, 7.3, 7.4, 7.5 and 7.6.

Generation, propagation and dissipation of transient events is a result of numerous elements, including network's layout and characteristics, transient sources, pipes and fluid's properties, etc. There are numerous causes of hydraulic instabilities, some inherently with a random nature (e.g. manual operational intervention). Similarly, many energy dissipation mechanisms, such as demand and leakage, acting to mitigate the amplitude and frequency of pressure fluctuations are pseudo-random (demand has a diurnal profile). Therefore, it is not unexpected that many pressure profiles contain random events of various magnitudes. However, sources related to the automated operation of the system (e.g. pumps and automated valves) generate pressure fluctuations that, in their proximity, are often repetitive (less random) and large (not fully dissipated yet).

Each group with its unique characteristics of pressure profiles and distribution of pressure fluctuation events impact the $CPIS$ metric (details in section 7.2). The metric is enhanced by considering the effect of significantly large isolated events and differentiating these from more frequent repeated pressure fluctuations. Further details on this is discussed in chapter 11.

7.1.1. Network assets and pressure fluctuations

Impacts of various network assets in generating pressure transient events have been researched extensively, e.g. (Wylie et al., 1993; Thorley, 2004). Examples and case studies, undertaken in this work, are presented here for three significant factors that affect the system instabilities, i.e. pumps, burst events and PRVs.

Switches pumps on and off are the source of a large proportion of pressure transient events. In Figure 7.6, an example of data collected at location downstream of a large pump station is presented. Labelled with triangle and circles, events are generated following a pump switch off and on respectively. The pump station comprises of three very large pumps with fairly complex

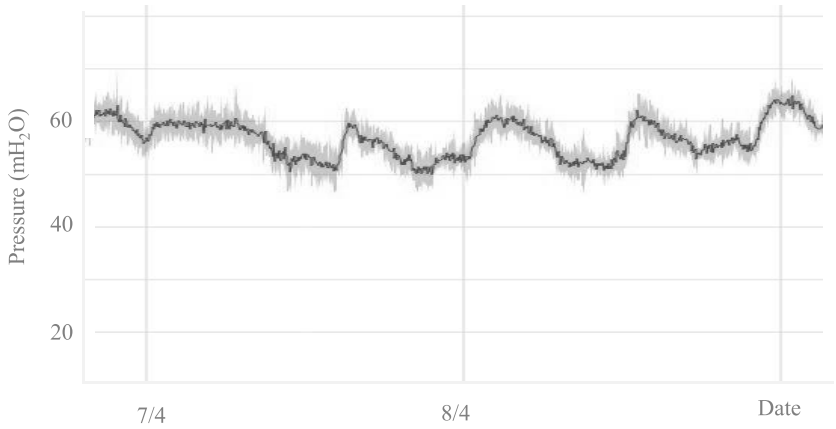


Figure 7.1.: Steady-state profile with moderate pressure variations (Group 2)

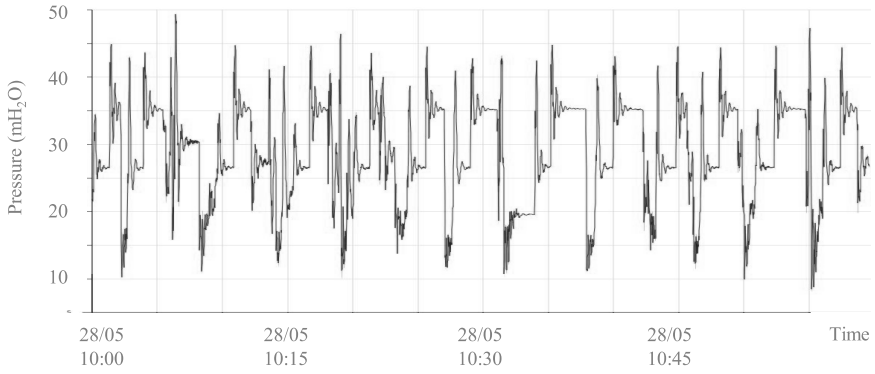


Figure 7.2.: Varying pressure profile with large pressure variation events (Group 3)

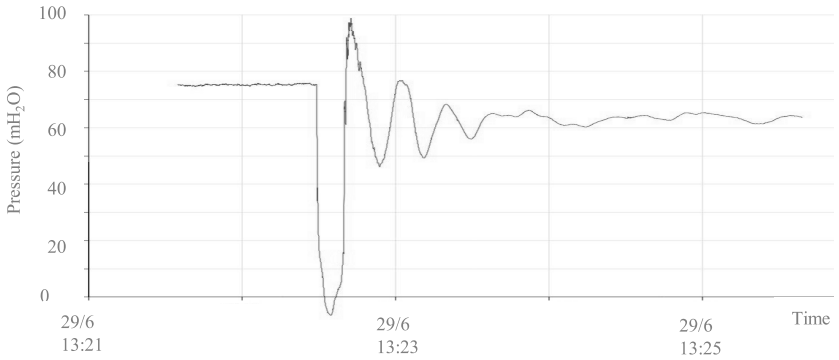


Figure 7.3.: A large transient event on a trunk main as result of value operation (Group 4)

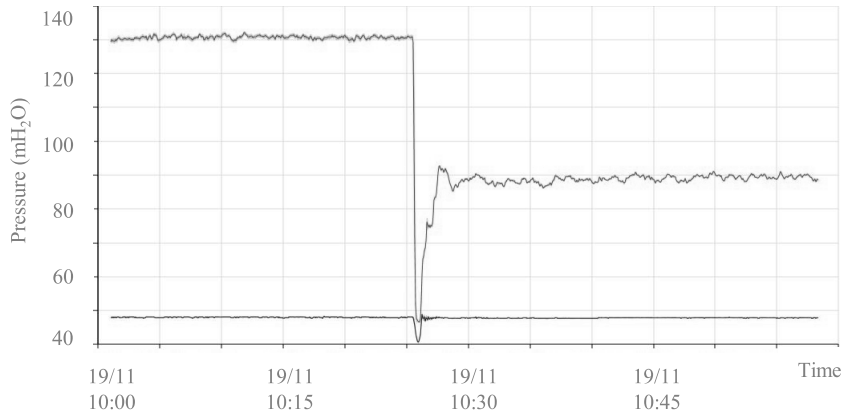


Figure 7.4.: Impact of a PRV in mitigating a pressure transient event generated upstream (upstream profile is Group 4, downstream profile is Group 1)

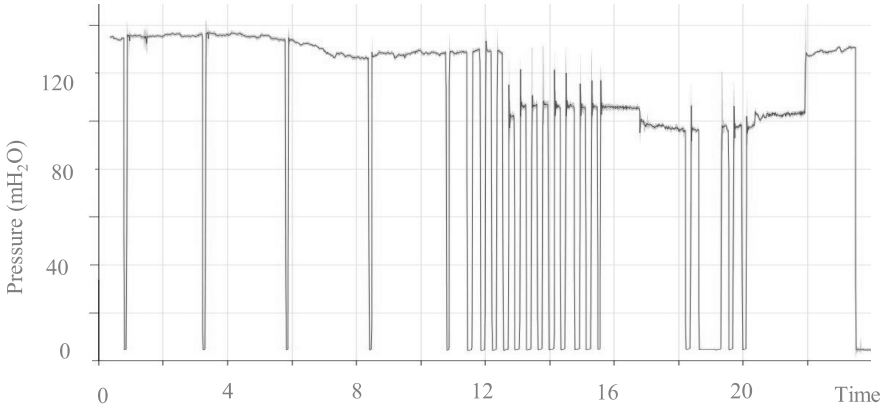


Figure 7.5.: Frequent and large pressure transient events (Group 5)

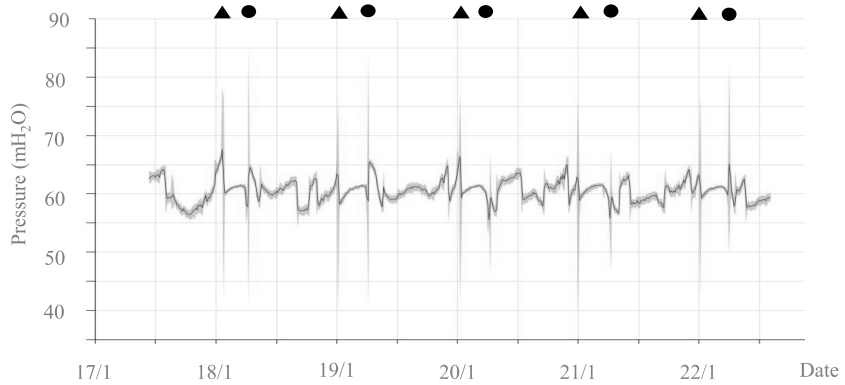


Figure 7.6.: Impact of pump switches in generating pressure transient events (Group 6)

operational settings and existing surge vessels. A pipe burst event occurred approximately 200m downstream of the pump station, resulting in significant water supply interruption. Root cause analysis was undertaken to investigate potential causes and contributing factors to the pipe failure. From the results of the pipe's laboratory chemical and physical testing and careful study of system operation and recent changes, the investigation concluded that pipe was in a good condition generally and the cause of failure was attributed to micro-cracks, intensified and propagated by repeated transient loadings, ultimately resulting in a pipe failure. The design and operation of the surge vessels had not been modified following various recent changes to the system and therefore no longer sufficient to mitigate all occurrences of transient waves effectively.

Pressure reducing valves are often configured to regulate the downstream pressure. Their impact on rapid pressure fluctuations are, however, not always clear. Presented in Figure 7.4, a PRV with fixed downstream head is shown to mitigate the upstream hydraulic instabilities, reducing an event size of approximately 48m to around 6m amplitude event. Review of other logging at PRVs suggest that, depending on PRV type and setting, and also characteristics of pressure transient event, PRV could act to dampen the amplitude of large events. In some other cases, pressure waves travelled through the PRV with minimum dissipating affect from the PRV. The energy from such events, when absorbed by the PRV, could contribute to deterioration of the PRV components and alteration in its settings. While some PRVs act to mitigate magnitude of transient events (e.g. Figure 7.4), oversize PRVs, when opening and closing extensively and in low flows, can themselves produce oscillations in downstream pressure profile.

7.2. Event extraction and cycle counting

Cumulative Pressure-Induced Stress (CPIS), as detailed in chapter 5, is calculated from equation 7.1 (introduced in section 5.2, repeated here):

$$CPIS = f[P_{mean}, DP] \tag{7.1}$$

A process, illustrated in Figure 7.7, is developed to calculate $CPIS$, for each dataset.

Diurnal steady state mean pressure (P_{mean}) is calculated from the hydraulic models for all locations. Dynamic pressure, DP, is a sum of amplitudes of

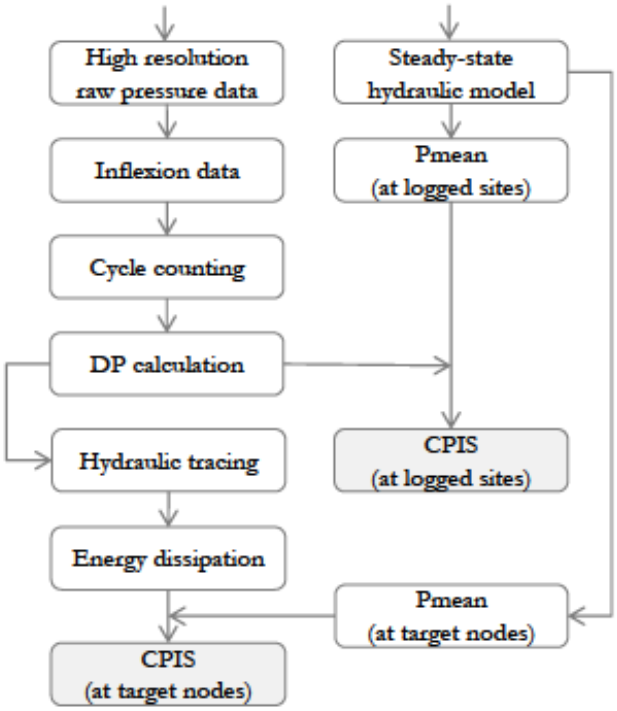


Figure 7.7.: Data analysis process

pressure cycles (σ_i) multiplied by frequency of each magnitude (n_i):

$$DP = \sum_{i=1}^{M} (\sigma_i \times n_i) \tag{7.2}$$

An inflexion coding and a cycle counting process, detailed below, are used to calculate DP.

7.2.1. Inflexion coding

The raw pressure data gathered in this project are series of pressure measurements in time intervals. The processes for data analysis, cycle counting and cumulative damage estimation (detailed in the following sections) require linear approximations of the time series data. There are several algorithms that transform the time series into series of straight line segments. However, most of these, often they have high computationally expense as they require multiple mathematical operations over the data. This is particularly relevant when the datasets contain large volume of data points. Collected pressure data is done at up to 128 samples/ second, to enable capturing rapid pressure fluctuation and pressure transient events. This results in a large number of data points for each dataset. Recording for a 24-hour logging period generates

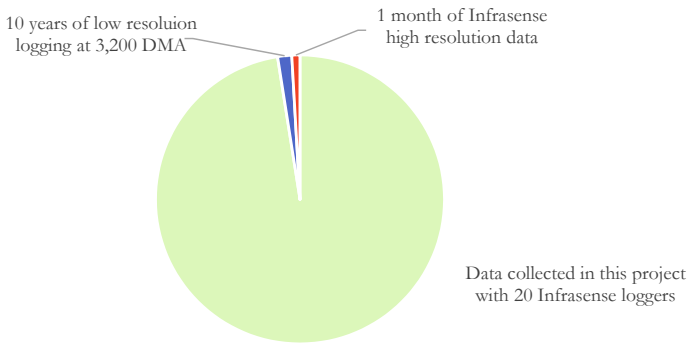


Figure 7.8.: Pressure database sizes using high and low resolution loggers

approximately 1.1×10^7 pressure measures. Figure 7.8 illustrates a comparison between the amount of pressure data collected in this project versus the data collected across the Company using 15-minute low resolution pressure loggers.

Hoskins (2015) developed a piecewise linear approximation method, *inflexion coding*, that requires only one pass over the data employing simple mathematical operation (addition, subtraction and comparison). The mains stages implemented are quantisation, extrema detection and step removal. Pressure data recorded in operational networks often contain background noise generated by system's hydraulic alteration and random demands. In the quantisation stage, noise in the raw pressure data, below a threshold, is removed and the waveform is converted to a series of steps. In the second stage, the waveform steps are simplified by removing the step points except the extrema, local minimum and maximum points. Finally, all steps are removed to produce a series of straight lines approximating the original waveform. Experimentally derived control parameters are introduced in order to preserve horizontal sections of the raw data (against deletion as part of step removal). Figure 7.9 illustrates an example of waveform and inflexion coding steps. Inflexion-coded data can represent the original raw data with fewer points, Figure 7.10, reducing computation complexity in the following processes.

7.2.2. Cycle counting

The next step in calculating DP is developing distributions of pressure cycle amplitudes and number of cycles for each magnitude.

To extract stress cycles from the pressure waveforms, there are various cycle counting techniques proposed in the literature. Cycle counting algorithms provide a distribution that contains amplitude of a pressure change and number of cycles for each magnitude. Lee et al. (2005) summarises different types consisting of one and two parameter methods. One-parameter cycle counting

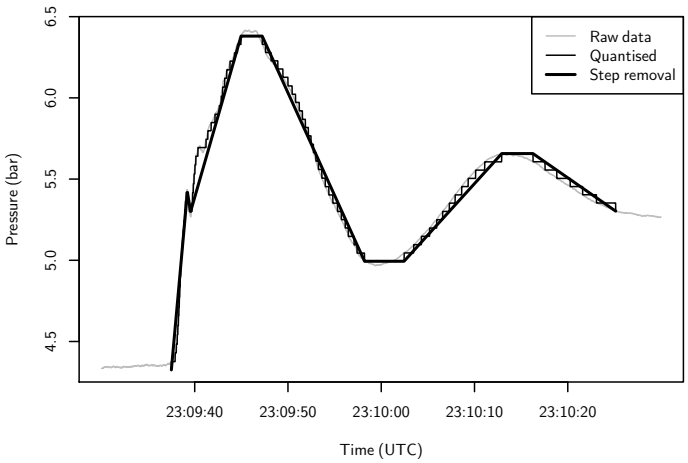


Figure 7.9.: Inflexion coding processing steps-from Hoskins (2015)

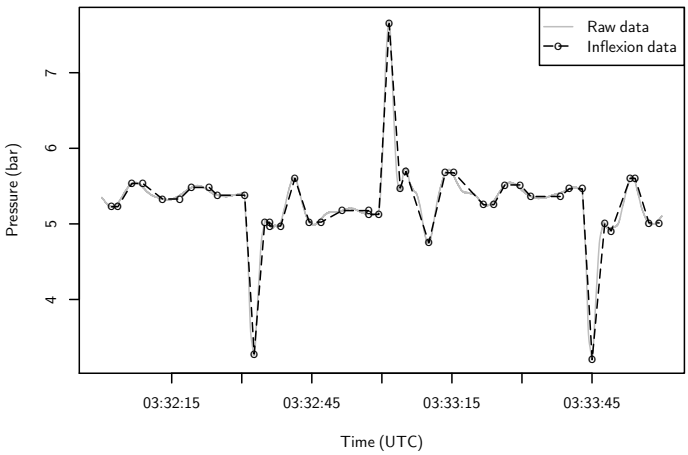


Figure 7.10.: Comparison of raw and inflexion-coded data-from Hoskins (2015)

method includes level crossing cycle counting, peak-valley cycle counting and range counting. Methods in two-parameter cycle counting include three-point and four-point cycle counting. Two-parameter cycle counting methods such as the *rainflow* cycle counting can represent variable-amplitude cycle loading.

Rainflow algorithm is the optimum method for cycle counting in fatigue deterioration studies, as highlighted in Rychlik (1993) and Kiefner et al. (1973). A recent study, (Hoskins, 2015), compared rainflow counting method with two other cycle counting techniques, half-cycle counting and maximum-edge counting, used in analyses of semiconductor systems. In half-cycle counting, each edge is treated as half a cycle (Figure 7.11a), while in maximum-edge counting, edges and counts of cycles with highest amplitudes are considered (Figure 7.11b).

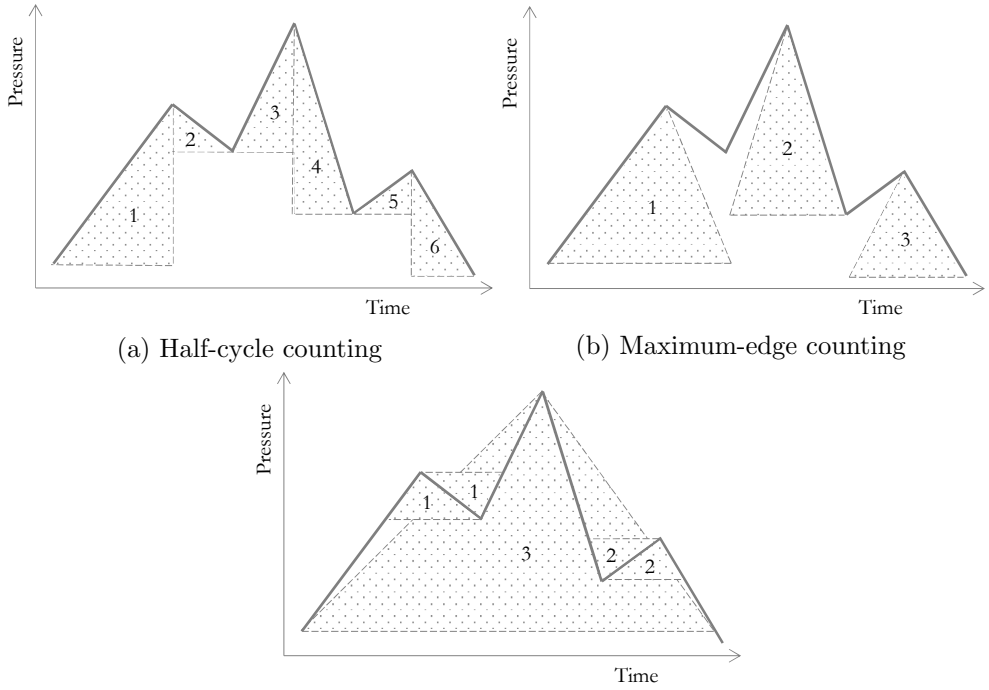
Rainflow counting method attempts to extract multiple stress cycles from each section of waveform (Figure 7.11c). The rainflow counting implementation method used here is the one developed by Downing and Socie (1982). This was adopted by American Society for Testing and Material (ASTM) as part of 'Standard Practices for Cycle Counting in Fatigue Analysis' in 1985 (Murakami, 1992), and in the re-approved version of the Standard in 2011 (ASTM, 2011). The algorithm is shown in Figure 7.12.

Hoskins (2015) simplified and improved the ASTM algorithm and restructured its operation using modern programming constructs, 7.13. In this algorithm, m and n are amplitude and number of cycle at that amplitude, C is frequency distribution, S starting point in the history, ϵ is a sequence which holds each new point as it is scanned, X is range under consideration, Y is previous range adjustment to X.

In comparing rainflow counting method with half-cycle and maximum-edge counting methods, Hoskins (2015) found variations in outcome metric generated by these methods, depending on the waveform's characteristics. Rainflow counting method captures more cycles in mid-amplitude ranges that are missed by other methods. For the analyses in this project, the results of rainflow cycle counting are used.

Table 7.1 illustrates an example for calculation of DP, as in equation 7.2, by employing rainflow cycle counting method. DP_i is normalised (DP_i/day) where the logging duration exceeds one day. Each row in the table represents an amplitude (e.g. cycle 1 or 2 in Figure 7.11c), with its frequency using rainflow cycle counting and the calculated normalised DP (in this case the logging duration 24 days). Normalised DP is calculated as summation of all values in last column of the table

Distribution of DP for assets logged is illustrated in Figure 7.14. Non-



(a) Half-cycle counting

(b) Maximum-edge counting

(c) Rainflow counting

Figure 7.11.: Cycle counting methods with cycles detected in each method, from Denk and Bakran (2013)

5.4.4 *Rainflow Counting:*

5.4.4.1 Rules for this method are as follows: let X denote range under consideration; Y, previous range adjacent to X; and S, starting point in the history.

(1) Read next peak or valley. If out of data, go to Step 6.

(2) If there are less than three points, go to Step 1. Form ranges X and Y using the three most recent peaks and valleys that have not been discarded.

(3) Compare the absolute values of ranges X and Y.

(*a*) If $X < Y$, go to Step 1.

(*b*) If $X \geq Y$, go to Step 4.

(4) If range Y contains the starting point S, go to Step 5; otherwise, count range Y as one cycle; discard the peak and valley of Y; and go to Step 2.

(5) Count range Y as one-half cycle; discard the first point (peak or valley) in range Y; move the starting point to the second point in range Y; and go to Step 2.

Figure 7.12.: Rainflow counting algorithm by ASTM ASTM (2011)

Process: IMPROVED RAINFLOW$(m_i \to \emptyset) \Rightarrow (\mathcal{C})$
Memory: $\mathcal{E} \leftarrow \emptyset$

if $m_i \neq \emptyset$ **then**
  $\mathcal{E} \leftarrow \mathcal{E} \cup m_i$
  while $|\mathcal{E}| \geq 3$ **do**
    $X = \left|\mathcal{E}_{|\mathcal{E}|} - \mathcal{E}_{|\mathcal{E}|-1}\right|$
    $Y = \left|\mathcal{E}_{|\mathcal{E}|-1} - \mathcal{E}_{|\mathcal{E}|-2}\right|$
    if $X \geq Y$ **then**
      if $|\mathcal{E}| = 3$ **then**
        $\mathcal{C} \leftarrow \mathcal{C} \cup \langle Y, 0.5\rangle$
        $\mathcal{E} \leftarrow \mathcal{E} \setminus \mathcal{E}_1$
      else
        $\mathcal{C} \leftarrow \mathcal{C} \cup \langle Y, 1\rangle$
        $\mathcal{E} \leftarrow \mathcal{E} \setminus \left\{\mathcal{E}_{|\mathcal{E}|-1}, \mathcal{E}_{|\mathcal{E}|-2}\right\}$
      end if
    else
      break ▷ Exit **while** loop
    end if
  end while
else
  Add each remaining edge in $\mathcal{E}$ as a *half* cycle to $\mathcal{C}$
end if

Figure 7.13.: Improved rainflow counting algorithm, from Hoskins (2015)

Table 7.1.: Example of DP calculation from rainflow cycle counting

Event's amplitude (mH_2O)	Event's frequency	DP_i	DP_i/day
...	...	...	...
0.340	720	245.04	10.21
0.459	912	418.48	17.44
0.516	3096	1598.12	66.59
0.655	288	188.72	7.86
0.769	14448	11112.28	463.01
0.825	2304	1901.51	79.23
...	...	...	...

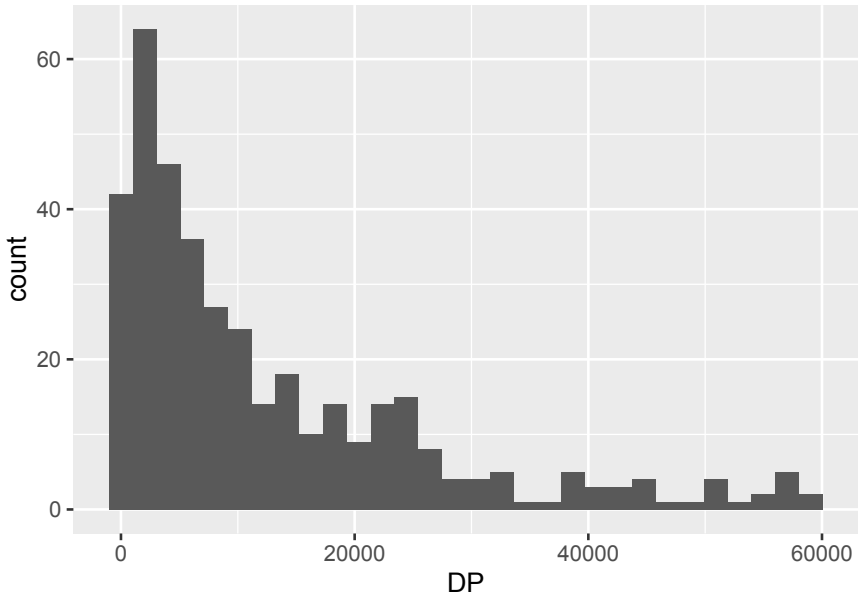


Figure 7.14.: Distribution of DP values for sites logged across the network

normality of the distribution is expected due to large number of sites having hydraulically calm operation or with few occurrences of hydraulic instabilities.

7.3. CPIS across the network

To study correlations between pipe burst events and impact of CPIS, this value needs to be estimated at the target locations (burst pipes or DMA inlets). As for a target location, where the pressure readings are not available from data collection, its CPIS is estimated based on the CPIS calculated at an upstream location.

Estimating CPIS at target locations require DP and P_{mean} values at these locations (equation 7.1). Steady state hydraulic models can be used to estimate P_{mean}. This could result in under-estimation, as the 15-minute interval pressure data often do not capture the actual pressure variability. However, in the absence of high resolution pressure data gathered on site, this is a viable alternative option. DP at target points (a measure of rapid pressure fluctuations), cannot be reliably estimated from steady state models. Instead, DP values can be estimated using collected high resolution pressure data. For this, hydraulic connectivity, hydraulic distance and energy dissipation mechanisms between the source node (where collected data is available) and the target node (where estimate for DP is made) need to be considered. Firstly, hydraulic connectivity and shortest path between nodes are determined. Then, energy

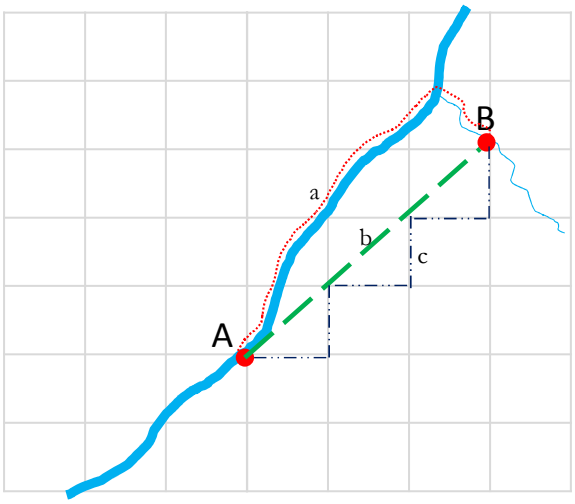


Figure 7.15.: Distance between network nodes using (a) hydraulic tracing method, (b) Euclidean, (c) Manhattan

dissipation mechanisms are analysed in order to estimate energy level at the downstream target nodes based on the energy level at upstream nodes and the dissipation. Details of these are discussed below.

7.4. Hydraulic tracing

There are several methods for estimating distance between nodes. Popular methods, such as Euclidean distance (straight-line distance between two points) or Manhattan distance (sum of the lengths of the projections of the line between the points onto the coordinate axes) do not produce reliable distance measures (Figure 7.15). The actual hydraulic distance between two points is driven by the physical characteristics of the network (i.e. pipe layout) and also presence of other assets altering system's hydraulic (e.g. closed valves, non-return valves, etc.).

Previous studies, such as Herrera et al. (2015), used graph theory as a measure of connectivity of each node to the source and to identify the shortest path between these nodes. This is done by calculating an index, weighted by the frictional resistance that the water needs to overcome to flow through the pipe; and the shorter path is the one with the lower resistance. The index takes into account total number of sources in the system and pipe length and diameter (as a surrogate measure for energy loss). The method is preliminarily developed for studying network's resilience by identifying the number of sources each node is connected to. It is, however, computationally expensive if all routes connecting nodes to sources are computed. This is specifically prohibitive, if

large networks of heavily populated nodes are to be analysed.

A simple yet reliable approach to check the hydraulic connectivity is by using steady state hydraulic models. An advantage in using hydraulic model for this purpose is that the analysis will not only check the hydraulic connectivity of two points, but it also provides information about other hydraulic relationship between the points. For instance, physical characteristics of the network and shortest path (pipe's length, material and diameter) and the network's control (such as demand and leakage) can be extracted from the model for each shortest path analysis. These information are used for analysing energy dissipation, as detailed in section 7.5. The main limitation of this method is availability of calibrated hydraulic models. However, hydraulic models, where not already available, are widely agreed as a fundamental component for network management, and many water utilities are in the process of developing such models.

7.4.1. Tracing techniques

Primarily developed to undertake water quality and water age studies, there are several tracing techniques available in standard hydraulic modelling tools, for instance Synergi modelling software (DNVGL, 2014). The main techniques are the following.

Supply trace For any node or facility in the model, the percentage of water came from one or more specified supply nodes can be estimated by supply tracing. Multiple supply nodes can be specified as starting points, but only supply nodes can be used as starting points (not other non-supply nodes). Results from this analysis are percentage of the water at a node or facility that entered the model through the specified supply node(s).

Backward trace A backward trace determines the percentage of water at a node or facility that will also flow through a designated starting node, i.e. the portions of the system that water passes through to arrive at a given node. For instance, if a given node j, the backward trace provides a number N_i for every node i in the system, such that $100 \times N_i$ provides a percent of water arriving at node j that passed through node i.

Forward trace With this analysis, for any node in the model, the percentage of water that came from a specified node is determined. Unlike a supply trace, it is not required to specify a supply node to perform a forward trace. Results of a forward trace are reported as the percentage of water at a node that also

flowed through the starting node. For instance, for any node j, a forward trace provides a number N_i for every node i in the system, such that $100 \times N_i$ is the percent of water at node i that passed through node j.

Supply trace algorithm requires the upstream nodes to be supply nodes (reservoirs, tanks, etc.). Many of the logging sites in this project are not at a source node and hence this method is not suitable. Vast majority of target nodes that need to be traced are downstream of logged sites. Therefore, backward tracing does not provide suitable framework for the tracing. Forward tracing is the most suitable method, as any node (in this case logged sites) can be assigned as supply nodes, providing tracing for other nodes downstream of these (target nodes). The supply contribution at node i from source s is calculated from equation 7.3 (DNVGL, 2014):

$$SC_{is} = \frac{Q_s + \sum_{i=1}^{N_l}(SC_{ls} \times Q_l)}{Q_i + \sum_{i=1}^{N_l} Q_l} \tag{7.3}$$

Where SC_{is} is fractional contribution of source s at node i, SC_{ls} is fractional contribution of source s in link l, Q_l is flow rate in link l, Q_s is supply node flow rate, N_l is number of links and Q_i is positive node flow at node i. Where the link l is fed by node i, SC_{ls} is equal to SC_{is}.

Calculation of supply contribution is based on the calculated flow rate and an *index array*. Index array directs the order in which the calculation at each node is performed. Calculation at each node is done only after all calculations upstream of the node are performed. The index array starts from a supply node (with only output flow) and moves downstream to each node when properties of all flows feeding the node are already calculated. For instance, as illustrated in Figure 7.16, tracing starts from node A (supply node), continuing to links 1 and 3. After this, the process continues to node B, as the only unknown for this node (link 1) is already calculated. Only then, it will move to node C (known links 3 and 4). The process continues along the links and nodes to cover the entire network. The index array calculated for this example is:

$$[A, 1, 3, B, 2, 4, C, 5, 6, E, 7, D]$$

Figure 7.17 illustrates a case study, CS1, an area comprising of 39 DMAs. A total of 40 site are logged using high speed pressure loggers, locations shown in Figure 7.17a. A forward tracing technique is used to analyse trace the flow in the network. The thematic map in Figure 7.17b illustrates the supply from logging site, FT501, and how flow from this node travels through other nodes and links in the network. That is the percentage of supply through each pipe

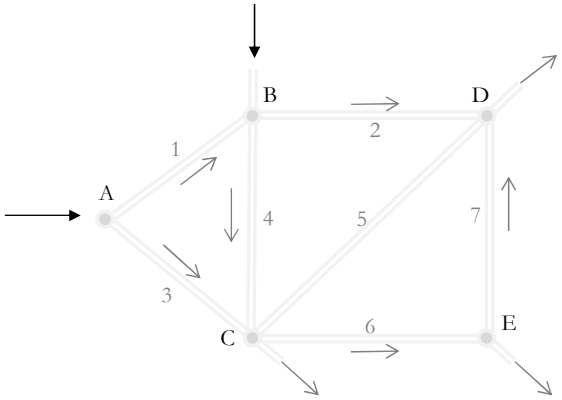


Figure 7.16.: Development of index array-adopted from DNVGL (2014)

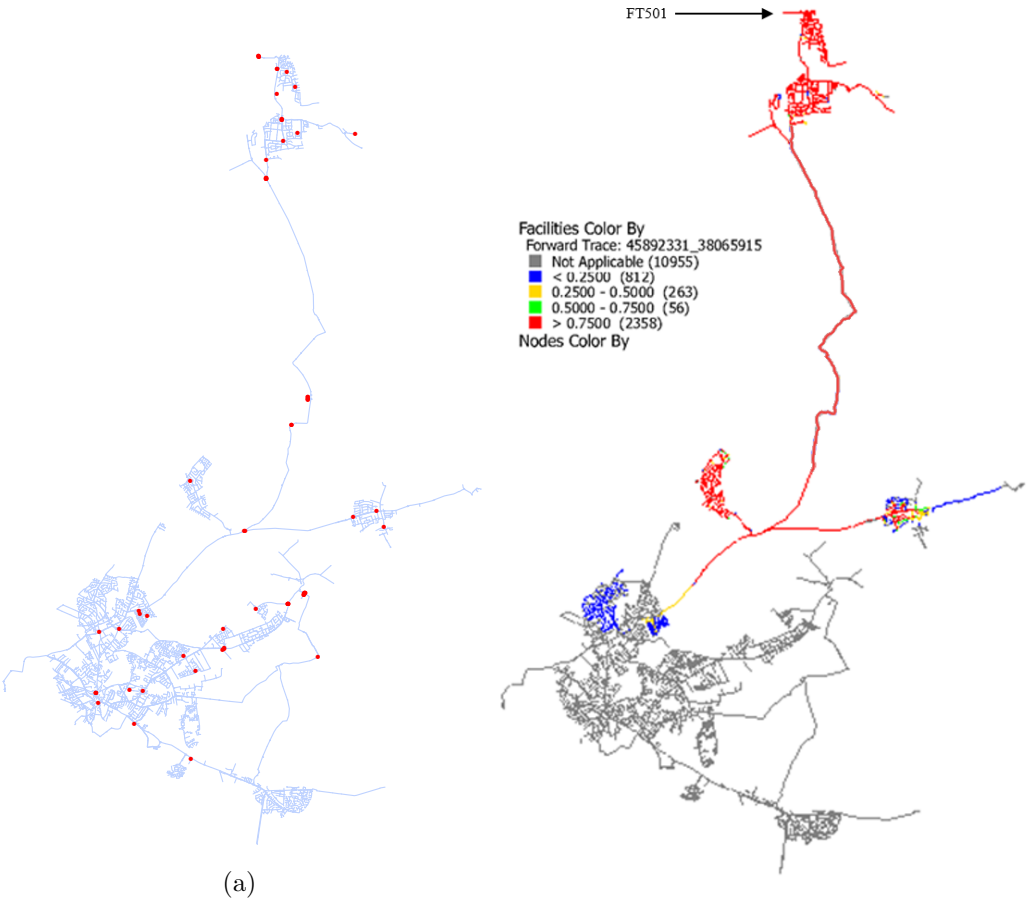
Table 7.2.: Forward tracing supply contribution from logging sites

Network nodes	FT501	FT515	FT516	FT517	...
...	...	...	...	...	
46014200_35780490	0.70%	0.70%	0.80%	0	
46131328_35827543	0.00%	0.00%	2.65%	0	
46139440_35826443	0.00%	0.00%	3.68%	0	
45899046_35778887	1.80%	1.80%	1.33%	0	
45901540_35778821	1.10%	1.10%	0.94%	0	
45905646_35778106	0.70%	0.70%	0.79%	0	
45908315_35778546	0.00%	0.00%	5.59%	0	
45913821_35779056	1.90%	1.90%	1.38%	0	
45913909_35779059	1.90%	1.90%	1.39%	0	
...	...	...	...	...	

in the network that has also passed through the supply node (FT501).

Individual hydraulic tracing run is undertaken for the entire network with every logging sites assigned as supply nodes in forward tracing algorithm and calculating the flow percentage for entire nodes in the network. A snapshot of the result is presented in Table 7.2. Each column is a logging site, set as supply node in each tracing run. Each row is a node in the network with corresponding value showing percentage of flow at this node that also passed through the source node. For instance, 0.8% of the flow at node 46014200_35780490 has also passed through FT516.

Forward tracing analysis provides the hydraulic connectivity and relationship between the source nodes and the downstream nodes inside the network. This is used to estimate CPIS value at network's nodes, based on CPIS at source nodes.



(a)

(b)

Figure 7.17.: (a) Source node (logging site FT501) (b) forward tracing analysis for FT501

7.5. Energy dissipation of hydraulic instabilities

There is a constant change in the hydraulic conditions of networks, due to frequent changes in flow demand, operation of assets such as valve and pumps and changes imposed by intervention by outside factors (e.g. manual valve operation), etc. Many of these hydraulic instabilities and transients decay rapidly, due to various conversion and dissipation mechanisms in the network. There has been much effort over the last few decades to study these mechanisms. Karney and Filion (2003) categorise various processes and identified six primary dissipation mechanisms, as detailed in section 2.6:

- Fluid friction
- Network demand
- Leakage
- Network characteristics
- Unsteady friction
- Specialised surge protection components

Investigation in the network's transient behaviour is conventionally done by analysing a pair of quasilinear hyperbolic partial differential equations (one-dimensional unsteady momentum and continuity equations):.

$$\frac{\partial Q}{\partial t} + gA\frac{\partial H}{\partial x} + RQ|Q| = 0 \qquad continuity \tag{7.4}$$

$$\frac{\partial H}{\partial t} + \frac{c^2}{gA}\frac{\partial Q}{\partial x} = 0 \qquad momentum \tag{7.5}$$

where A is the cross sectional area, g is gravity acceleration, c is the wave propagation speed, R is frictional resistance, H is piezometric head and Q is flow. The continuity equation, represents the effects of fluid compressibility and conduit elasticity, and the momentum equation represents the inertial effects of fluid acceleration and deceleration (Karney et al., 2014).

Energy dissipation during transient events have received considerable attention over last few decades. Most studies primarily consider alterations to the momentum equation to account for unsteady friction effects (Karney et al., 2014).

Using what is known as *quasi-steady approximation*, friction losses during the transient pipe flow simulation are estimated based on steady-state flow

conditions. As described by Silva-Araya and Chaudhry (1997), this is based on the main assumption that the transient conditions headloss is equal to the headloss for steady uniform flow where an average velocity is equal to the instantaneous transient velocity. This assumption is, however, not always accurate as velocity profile is not necessarily the same as uniform flow profile. Furthermore, pipe shear stress in transient condition and mean velocity are not in phase, resulting in calculated friction losses in transient flow to be inaccurate (Wylie et al., 1993).

Much work has been undertaken in recent decades in developing unsteady pipe flow friction models in the place of steady friction models (such as Darcy-Weisbach and Hazen-Williams). Karney and Filion (2003) lists several of the one-dimensional models developed to explicitly consider boundary-layer effects and unsteady shear stress between fluid and conduit wall during transient events. For instance, Silva-Araya and Chaudhry (1997) develops a dimensionless energy dissipation factor, the ratio of energy dissipation in transient flow to energy dissipation in quasi-steady conditions that quantifies the amplitude of the energy dissipated in excess of the quasi-steady formulation. The factor is used to adjust friction term in the momentum equation.

Two important factors that should be considered to enhance performance of numerical models concerning transient energy in the network are transient effects on the wall shear stress and the pipe material visco-elasticity (Meniconi et al., 2014). Several investigations, such as Brunone et al. (2000), Covas et al. (2005) and Duan et al. (2010), studied the effect of conduit visco-elasticity on unsteady energy dissipation. More recently, Meniconi et al. (2014) examined the effect of a partially closed in-line valve, visco-elasticity and unsteady friction on the transient behaviour of a pressurised pipe. In their study, Covas et al. (2005) found a good agreement between the pressure variations obtained with the linear visco-elastic model and experimental data. Furthermore, the linearised visco-elastic contribution on pressure peak damping and phase shift was established to be more significant than contribution of the unsteady friction effect.

Considering an energy-based approach, Duan et al. (2010) and Meniconi et al. (2014) illustrated that understanding of network's transient behaviour can be enhanced by considering energy transformations. Manipulating the standard equations of transient pipe flow presented in Chaudhry (1987) and Wylie (1978). Karney (1990) developed an energy expression that accounts for the internal and kinetic energy of the fluid as well as work and mechanical-dissipation processes. The energy expression for transient closed-conduit flow in a single pipe is:

$$\frac{\rho A}{2}\left(\frac{g}{c}\right)^2\frac{d}{d_t}\int H^2 d_x + \frac{\rho A}{2}\frac{d}{d_t}\int V^2 d_x + \frac{f\rho A}{2D}\int |V|^3 d_x + \rho g A V(L,t)H(L,t) - \rho g A V(0,t)H(0,t) = 0 \quad (7.6)$$

where D is pipeline diameter, V is average fluid velocity, f is Darcy-Weisbach friction factor, ρ is the density of water (1000 kg/m^3), c is celerity of the shock wave, g is acceleration due to gravity, $V(L,t)$ and $H(L,t)$ are the velocity and head at $x = L$ and $V(0,t)$ and $H(0,t)$ are the velocity and head at $x = 0$. In equation 7.6, the first term is the rate of change of internal energy. The second and third terms denote the rate of change of kinetic energy and the rate of viscous dissipation in the system. The fourth term denotes the thermodynamic work performed at the ends of the pipeline.

Karney and Filion (2003) used this energy expression to investigate the role of steady friction, unwanted leaks, topological complexity and surge control devices in the decay of transient energy. It was found computationally straightforward as it make use of head and velocity terms already calculated by a hydraulic solver. Some of the limitations of this expression, as highlighted by Karney and Filion (2003), are that the current form of the expression does not cover dissipations caused by vibrations, trapped air and hysteresis effects in the elastic behaviour of pipe walls and confining soil.

7.5.1. Limitations

Most one-dimensional models of transient energy study, listed in Karney and Filion (2003), run a simple reservoir-pipe-valve system. Simulations developed by Karney and Filion (2003), run on a looped and symmetrical network with some topological complexities. Developing and performing simulations of transient energy dissipation for large, often complex networks consisting of multiple operational assets, sources, complex topologies and surge-mitigation devices is a big challenge. Adding to this, lack of reliable data about buried assets and high resolution data on the system hydraulic, and the challenge can prove extremely difficult.

7.5.2. Simplified methods

In order to estimate CPIS at each target node in the network, energy dissipation mechanisms need to be studied to estimate the energy level (here amplitude of pressure variations cycles) at each node. However, existing models and methods fail to provide capabilities for such analysis in large complex networks.

Furthermore, considering the scale of the current study, detailed investigation and analysis on individual pipes and assets are not practical.

Three simplified methods are proposed and reviewed in this work in order to estimate DP for each node in the network based on energy dissipation between an upstream node (logged location) and a downstream node (target location).

- *Method 1:* Flow weighting method
- *Method 2:* Flow weighting method plus static headloss
- *Method 3:* Energy dissipation linear model

In these methods, it is necessary to identify the source of hydraulic instabilities. Then, sub-system analysis and hydraulic tracing are used to identify sections of the network that are impacted by each source of hydraulic instabilities. For areas impacted, the energy levels are estimated using these methods.

For the instabilities generated by the operational assets in the network (these cover most instability cases), it is practical, and with relatively small level of error, to identify the sources of instabilities. In other cases, where random events, e.g. human intervention or a pipe burst event, cause instabilities, it is not possible to reliably associate transient events to a particular source, without detail investigation. Several methods exist in the literature to identify the source of pressure transient events. However, these often require deep knowledge of the network characteristics and hydraulics and demand detailed analysis. For the purpose of this work, it is assumed that the dominant sources of hydraulic instabilities are the operational assets (pumps, automated valves, etc.). This is however, a limitation of this approach and where possible, a detailed investigation of transient sources (random and non-random) should mitigate potential errors. Further details on each method is described below.

7.5.3. Flow weighting method

In the simplistic method, the estimate for energy level at each target nodes is made based on the level of water flow reaching the node that also passed through source node. In this method, DP at each node is calculated based on DP at the source node using the flow weighting at each node. The flow weighting is obtained from the forward tracing analysis. Here, the assumption is made that the DP distribution and mitigation is linearly proportionate to the level of flow distribution between the source and the target nodes. Where there is more than one source node providing flow to a downstream target

nodes, DP at target nodes is calculated proportionally based on the flow weighting the node receives from each of the multiple source nodes.

This basic approach ensures the hydraulic connectivity between the source and downstream nodes, and that flow proportion is an indication of proximity and flow and headlosses. This is however a simplistic approach and has several limitations. Network physical characteristics, layout and pipes properties impacting friction and energy waves transformation and propagation are not explicitly considered. Furthermore, various energy dissipation mechanisms, such as leakage, demands and unsteady friction are also ignored.

7.5.4. Flow weighting method plus static headloss

To consider network and pipe physical and hydraulic characteristics, method 1 is enhanced considering fluid steady state friction, as an important energy dissipation mechanisms. In this method, the same procedure as method 1 is undertaken, with an additional step to consider the steady state headloss between the source and target nodes. The amplitudes of the pressure cycles that form the DP calculation at source node are moderated (reduced) by the headloss from this node to each target node. The process involves calculation of headloss between the sources and each target nodes in the network. This is obtained by running steady state hydraulic simulations. Maximum level of headloss in a 24-hour simulation run is used, as a conservative measure to ensure energy levels at target nodes are under-estimated, rather than over-estimated. Similar to method 1, the final DP will then be calculated by accumulating the DP's from each source node. If more than one source feeds a downstream target node, weighted factors are used based on the source flow weighting from the tracing analysis. An schematic illustration of this is presented in Figure 7.18, with overview of processes to calculate DP in Figure 7.19.

Compared to method 1, this approach takes into account friction headlosses, an important energy dissipation mechanism. This is also a surrogate measure for some elements of the network characteristics and hydraulic behaviour. However, the approach has limitations in explicitly considering energy dissipation as result of leakage and demand, as well as unsteady friction losses and consideration on pipe-fluid interactions and energy transformation.

7.5.5. Energy dissipation linear model

In order to address the limitations with method 1 and 2, a further method is developed. This method takes a statistical approach to investigate the relationship between the energy level at various nodes across the network.

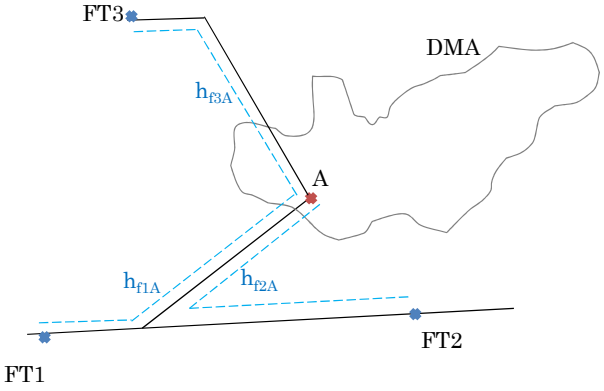


Figure 7.18.: Headloss when calculating DP at node A (*Method 2*)

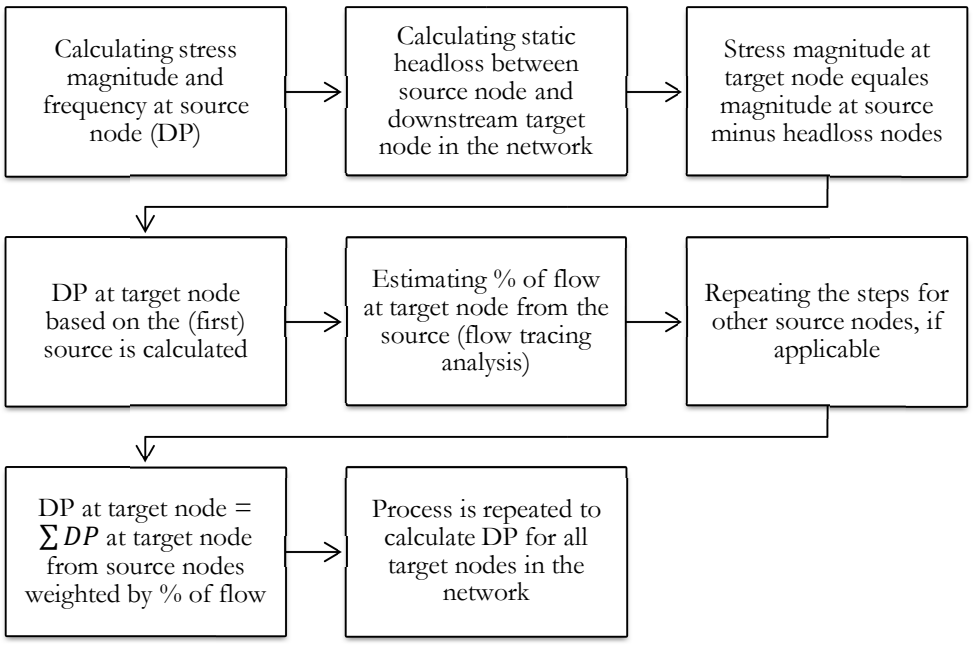


Figure 7.19.: Algorithm to estimate DP at each target node using *Method 2*

Factors representing the main energy dissipation mechanisms are considered in developing a statistical model that can then be used to estimate the energy levels at other locations. High resolution pressure data is gathered across the network to identify the hydraulic instabilities (and the DP representing the energy level) at each node.

A total of 27 locations are logged in a case study area, CS1. Logged sites are designed to include locations on or just downstream of operational sites (as source nodes closest to operational assets generating the hydraulic instabilities), locations on trunk mains, DMA inlets as well as nodes on smaller distribution mains inside DMAs (as target nodes receiving the energy). DP is calculated at each logged site, and a multi-linear statistical model is developed to investigate impact of various factors in dissipation of DP in the network.

These factors represent the energy dissipation mechanisms include hydraulic distance between nodes (shortest path), characteristics of pipes between nodes (material and diameter), number of customers and demand volume between the source and target nodes. The variables used in the model are developed as two cohorts of pipe diameter (250mm as threshold), two groups of pipe material (metal and non-metal) and demand volume and number of customers on the identified shortest path.

Hydraulic models and forward tracing algorithms are used to obtain these factors for every combination between the logged locations at source and target nodes. Figure 7.20 illustrates en example of pipes selected between two logged sites. Characteristics of the highlighted length of pipe, that might be comprised of a combination of different pipe characteristics, are used to calculate the factor in the model, as presented in Table 7.3. For instance, from source A to target node B, 92.8% of shortest path between the nodes is metal pipe and the entire length is pipes larger than 250mm, with 6 customers fed from the link and demand volume of 39.9 m^3/day. It should be noted that some of these links are feeding other DMA in addition to the number of customers directly fed from the pipe.

The combination of source and target nodes, using the 27 sites logged, created a dataset with 67 observations used in model development. 2/3 of the dataset (44 observations) are used to develop the model with remaining 1/3 used to validate the model.

Model run and output The model, presented in equation 7.7, is developed by fitting a multi-variate linear model to the observation regarding DP at the source and target nodes and characteristics of the network between these nodes.

Table 7.3.: Factors for DP calculation using multi-variate linear model

Route ID	DP (source)	DP (target)	Metal length (%)	$\leq 250mm$ length (%)	Demand volume (m^3/day)	Number of customers
A to B	575.3	405.08	92.8	0	39.9	6
D to C	575.3	1015.1	38.8	23.5	17	4
E to F	267.8	525.9	44.1	84.8	18.5	56
A to C	1037.6	4875.2	19.6	100	11.1	26
G to F	1037.6	889.9	100	17.2	23.0	45
...	...	...	...	...	...	...

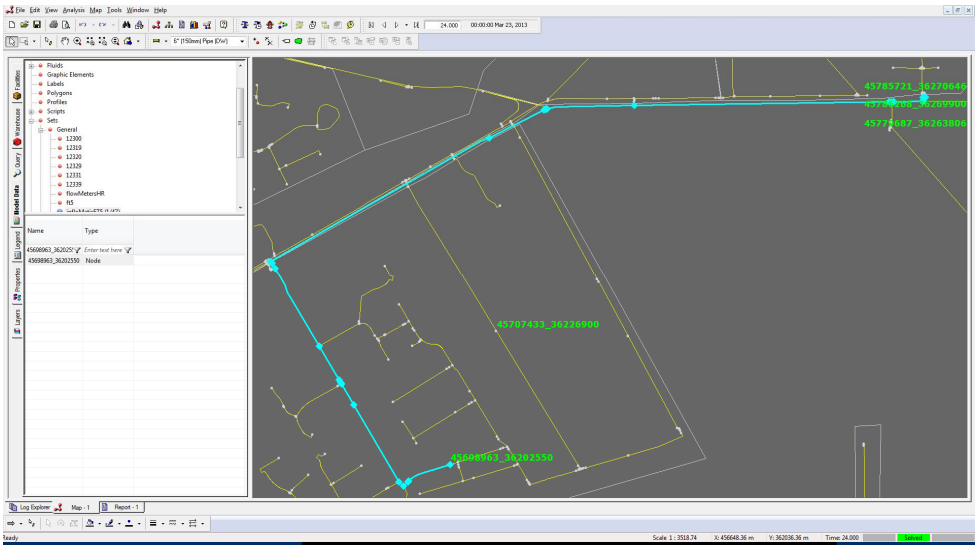


Figure 7.20.: Example of calculating hydraulic shortest path and pipes characteristics (screen-capture from simulation run

$$\log(\frac{DP_A}{DP_B}) = -2.545 + 1.779 \times (MetalLength) + 1.207 \times (\leq 250mmLength) \\ -0.0034 \times (DemandVolume) + 0.0045 \times (NumberofCustomers) \tag{7.7}$$

As illustrated in equation 7.7, material type and pipe diameter play important part in energy dissipation (DP) between two nodes. Significantly smaller coefficients for demand volume and number of customers suggest lesser impacts from these factors. This does not align with the notion that highlights importance of these factors in energy dissipation. The random nature of demand and relatively small number of customers on pipes in the observations potentially contribute to this uncertainty. There are also other limitations in this approach. While pipe properties are representing the network and hydraulic characteristics, the explicit consideration of the energy dissipation mechanisms factors are not included in the model. Furthermore, estimations for level of leakage are often prone to uncertainties and only available at area levels (e.g. DMA) but not pipe level. This prevents representing of the leakage factors in the investigations, specifically those studying large networks, as oppose to a simple pipe system.

Using equation 7.7, the values for DP at any target point can be calculated based on DP of a upstream (source) node and characteristics of the network between the nodes. This method is used to estimate DP at downstream nodes in Case Study 1, as described in chapter 8, for the analysis of correlation in the statistical models.

8. Multi-variate analysis of pipe failures

In this chapter, a multi-variate analysis of pipe failure records is carried out and impact of various predictor variables is investigated. The study set out with the aim of assessing the effect of dynamic pressure variations on pipe degradation and ultimate failure.

A multiple logistic regression modelling approach is applied and several models are developed to study the relationship between predicted probability of pipe failure and various factors that can impact the pipe deterioration.

Three case studies and their associated datasets are considered for the model development. Overview of the case studies is presented in section 8.3. Physical and hydraulic characteristics of the case studies, including single-variate analysis of historical pipe failure records in these areas are described in chapter 4. Four models are fitted to these datasets and tested for statistical significance, goodness of fit and predicted values.

In this chapter, firstly, an overview of the multiple logistic regression concept is provided. Then, various statistical tests that are used to assess the models are described. Three case studies used for the modelling are described in the following sections. After that, model development stages are explained with step by step development of one of the models. At the end, list of final fitted models is presented. This chapter focuses on model development and testing, while the results and discussions are provided in chapter 9.

8.1. Multiple logistic regression

Multiple logistic regression is used to explore associations between one (dichotomous) outcome variable and two or more predictor variables that may be continuous, ordinal or categorical. This modelling approach is particularly suitable for this study, as the outcome variable, pipe failure, is a dichotomous outcome with outcome *success* defined as pipe failure.

The purpose of multiple logistic regression is to isolate the relationship between the predictor variables and the outcome variable from the effects of

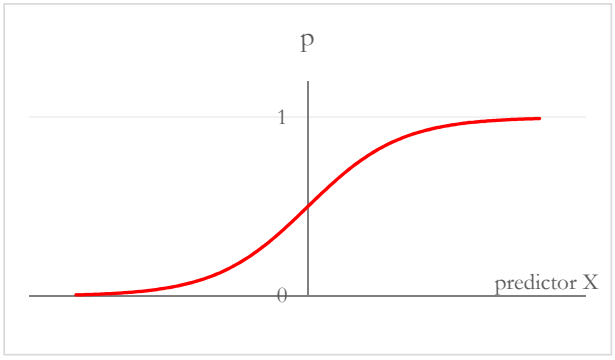


Figure 8.1.: Logistic probability curve

one or more other variables (confounders).

There are several advantages in using the logistic regression modelling approach. These include the flexibility and relative ease of calculations mathematically as well as ease of interpretations. This also allows for non-linear response, where, following changes in linear predictor variable, there would be minimum impact on the outcome variable, specifically where probability p tends to 0 or 1 (Figure 8.1).

Logistic regression does not make many of the key assumptions of linear regression and general linear models that are based on ordinary least squares algorithms. It does not need a linear relationship between the dependent and independent variables. Logistic regression can handle all sorts of relationships, because it applies a non-linear log transformation to the predicted Odds Ratio (detailed below). Moreover, the independent variables do not need to be multi-variate normal, and the residuals do not need to be multi-variate normally distributed. Logistic regression can handle ordinal and nominal data as independent variables.

The general formula for the logistic probability model is defined in equation 8.1, where $\mathbf{X} = (X_1, ..., X_n)$ represents predictor variable(s) and β is model coefficient(s), and $p(X)$ is probability of $Y = 1$:

$$E(Y|X) = p(X) = \frac{1}{1 + e^{-g(X)}} \tag{8.1}$$

where

$$g(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n \tag{8.2}$$

and

$$Y = E(Y|X) + \epsilon \tag{8.3}$$

ϵ, representing error, follows a shifted Bernoulli distribution.

Logit transformation of g(X), equation 8.4, is desirable as it provides properties of linear regression model, i.e. it is linear, continuous and ranges from $-\infty$ to $+\infty$.

$$g(X) = \ln(\frac{p(X)}{1-p(X)}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n \tag{8.4}$$

Maximum likelihood is the general method of estimation that yields values for unknown parameters that maximise probability of obtaining observed data. Likelihood function is probability of observed data as function of unknown parameter. Assuming observed data are independent, likelihood function is (Hosmer et al., 2013):

$$l(\beta) = \prod_{i=1}^{n} p(X_i)^{y_i}[1-p(X_i)]^{1-y_i} \tag{8.5}$$

Working of the logistic regression model is to choose coefficients, β_i, so that $l(\beta)$ is maximised. It is methodically easier to work with the log of this:

$$L(\beta) = \ln(l(\beta)) = \sum_{i=1}^{n} \{y_i \ln(p(X_i)) + (1-y_i)\ln(1-p(X_i))\} \tag{8.6}$$

Where to get β_i that maximise $L(\beta)$:

$$\begin{cases} \dfrac{\partial L(\beta)}{\partial \beta_0} = 0 \\ \dfrac{\partial L(\beta)}{\partial \beta_1} = 0 \end{cases} \tag{8.7}$$

Resulting in likelihood equations:

$$\begin{cases} \displaystyle\sum_{i=1}^{n} [y_i - p(X_i)] = 0 \\ \displaystyle\sum_{i=1}^{n} X_i[y_i - p(X_i)] = 0 \end{cases} \tag{8.8}$$

For logistic regression, these are non-linear and therefore an iterative procedure is required to solve. Iterative weighted least square procedure is seen as

one of the best method to do this.

8.1.1. Odds ratio

As calculations in the logistic regression model use logarithmic values, interpretation of the result might not be intuitive. *Odds Ratio* (OR) test is used as a more intuitive way of describing the change in probability of outcome (here pipe failure) with changes in the predictor variable. Linear function of predictor variables, g(X), as defined in equation 8.4,is:

$$g(X) = \ln \frac{p(X)}{1 - p(X)} = \ln(\frac{\text{probability of outcome occuring}}{\text{probability of outcome not occuring}}) \tag{8.9}$$

The OR is as the exponential of the difference between the two logit transformation functions. Where the predictor variable is binary, the OR is $exp(\beta_i)$, where β_i is the predictor's estimated coefficient.

It should be noted that *odds* of an event is the ratio of number of occurrence of an event to number of time the event does not happen, while the event's *probability* is the ratio of the number of time event happens to the number of times it could happen.

8.2. Model statistical tests

In order to evaluate the model fit and performance, several statistical tests can be undertaken. Here, tests on statistical significance of predictor variables (Wald test), model goodness-of-fit (Likelihood ratio test and McFadden pseudo-R^2), and validation of predicted values (ROC curve), are undertaken.

8.2.1. Statistical significance

Wald test The objective of significance testing is to establish whether each of the predictor variables within the model contribute to the relationship between predictor variables and the outcome variable. The Wald Test is used to determine how close an estimate of a coefficient within a statistical model is to the true value of that coefficient. This defines significance of each predictor variable. In the process of model development, the predictor variables identified as not being significant are removed from the model. Dataset will then be refitted to the probability model testing other predictor variables.

Wald test, for a normal distribution, is calculated as:

$$Z = \frac{\beta - \beta_0}{SE} \tag{8.10}$$

β is the maximum likelihood estimate of the coefficient, β_0 is the value of the parameter that is being tested and SE is the standard error in β.

Here, the null hypothesis is that the coefficient of each model's predictor is significantly different from zero. If it fails to reject the null hypothesis, i.e. Pr<0.05 at conventional 95% level, this suggests that removing the predictor would not impact the model fit substantially. Values from the Wald tests are presented in the models' coefficent estimate tables, e.g. Table 8.4, as $\Pr(>|z|)$.

8.2.2. Goodness-of-fit

Goodness-of-fit investigate whether predicted values are accurate representation of observed values.

Likelihood ratio test In studying the model significance, the proposed model is compared to the null and saturated models. In the *saturated* model, it is assumed that each data point from the total points of n has its own parameters, hence there are n parameters to be estimated. The *null* model assumes only one parameter for all the data points, resulting in only one parameter to be estimated. The *proposed* model assumes that the data can be explained with m parameters and one intercept term, resulting in $m + 1$ parameters. If LL = log likelihood, then:

$$\text{Null Deviance} = 2(LL(\text{Saturated Model}) - LL(\text{Null Model})) \tag{8.11}$$

on DOF being the difference between the DOF of the saturated model and DOF of the null model (Hosmer et al., 2013). Similarly:

$$\text{Residual Deviance} = 2(LL(\text{Saturated Model}) - LL(\text{Proposed Model})) \tag{8.12}$$

on DOF being the difference between the DOF of the saturated model and DOF of the proposed model.

Residuals represent the difference between the data and the model. Null deviance indicates the response predicted by a null model. Residual deviance

represents the response predicted by a proposed model adding predictor variables, indicating the contribution of each observation to the overall deviance; the lower the value, the better the model. The residual deviance can be used to test whether the null hypothesis is true.

Deviance, D, can be defined as:

$$D = -2 \ln \frac{LL_k}{LL_s} \tag{8.13}$$

where LL_k is the log likelihood of the fitted model with k variables and LL_s is the log likelihood of the fully saturated model. The statistic D follows a chi-square distribution with the number of degrees of freedom (DOF) equal to the difference in number of variables between the two models. A saturated model is one that contains as many parameters as data point.

D, in logistic regression is the same as residual sum of sequence in linear regression; that is sum of squared deviation of observed data about the regression line. When model outcome is binary, the likelihood of saturated model is equal to 1; and therefore:

$$D = -2 \ln(\text{likelihood of fitted model}) \tag{8.14}$$

The difference between two deviance values of D_{null}, deviance for the model without variable, and D_k, deviance for model with k variables, is often referred to as G for goodness of fit.G is often used to compare the fit of the model with and without the predictor variables (Cohen et al., 2013):

$$G = -2\ln \frac{\text{likelihood without variable}}{\text{likelihood with k variables}} = -2 \ln \frac{l_{null}}{l_k} \tag{8.15}$$

The difference between the null deviance and the model's deviance follows a chi-squared random variable with k DOF, equal to the difference between the null model's DOF and the proposed model's DOF. In order to test for significance, p can be calculated as following:

$$p = 1 - pChiSq(\text{deviance, DOF.residual}) \tag{8.16}$$

It should be noted that p is not p-value of the chi-squared distribution, but 1 minus that value, as presented in equation 8.16. If $p > 0.05$, then the model residual's deviance on residual's DOF is not significant at the conventional five percent level, so there is not sufficient evidence against this model to reject it. If $p < 0.05$, then this is significant, and an improved model is required.

McFadden pseudo-R^2 In order to investigate predictive power of the logistic regression model, several pseudo-R^2 metrics have been proposed in the literature, the most notable of which is McFadden's R^2. This is defined as:

$$1 - \frac{LL_f}{LL_{null}} \tag{8.17}$$

where LL_f is the log likelihood value for the fitted model and LL_{null} is the log likelihood for the null model. McFadden's pseudo-R^2 ranges from 0 to just under 1, where values closer to zero indicating that the model has no predictive power.

McFadden pseudo-R^2 can be considered from approach of explaining the variability, similar to the conventional R^2. In this case, the log likelihood of the null model, in McFadden pseudo-R^2, is treated as a total sum of squares. Similarly, the log likelihood of the full model is treated as the sum of squared errors. Furthermore, the ratio of the likelihoods suggests the level of improvement over the null model offered by the full model (Long and Freese, 2006).

McFadden firstly suggested the pseudo-R^2 in McFadden (1973) and later in Zarembka (1974). In his book (McFadden, 1977), McFadden discusses the relationship between pseudo-R^2 and conventional R^2, showing that rate of increase in pseudo-R^2 is slower than R^2. He argues that the values for his pseudo-R^2 index tends to be considerably lower than those of the R^2 index, suggesting that, for example, values of 0.2 to 0.4 for pseudo-R^2 represent *excellent* fit. Simulations by Domencich and McFadden (1975) found that the 0.2-0.4 range for pseudo-R^2 to be equivalent to 0.7-0.9 range obtained for R^2 for a linear function. Other researchers, such as Louviere et al. (2000), also stated that values of pseudo-R^2 between 0.2 and 0.4 are considered to be indicative of extremely good model fits.

8.2.3. Validation of predicted values

ROC curve The Receiving Operating Characteristic (ROC) is a measure of classifier performance. The curve is developed by plotting the true positive rate (positive data points that are correctly considered as positive) against the false positive rate (negative data points that are mistakenly considered as positive) at various cut-off point. The true positive rate and false positive rate are also referred to as *sensitivity*, and *1-specificity*, respectively. The curve illustrates a trade-off between sensitivity and specificity. The closer the curve to the left and top border of the ROC space, the more accurate the model. The slope of the tangent line at each point provides the likelihood

ratio of the point value of the model. The area under the curve , or AUROC, ranges from 0.5 to 1.0. An AUROC closer to the value 1 represents a better model in discriminating between the two categories which comprise the target variable. There is no specific threshold or cut-off point, and often new studies compare their AUROC against the similar studies. Studies in medicine have considered AUROC values larger than 0.7 to be 'fair' and values larger than 0.8 to be 'good' (Tape, 2016). In applied psychology, values of larger than 0.7 are considered strong effect (Rice and Harris, 2005). For this study, value of 0.7 is considered to be the threshold for AUROC.

8.3. Model case studies and datasets

Three case studies are considered for model development and analysis, as described in chapter 3 and repeated below.

Case study 1 (CS1) This is an area comprising of 39 DMAs, with total pipe length of 215.04 km. Total number of pipe failure records (outcome variables) in model development is 242. Higher density of logger deployment was used for high frequency pressure data collection in this area. Moreover, hydraulic tracing methodology (detailed in chapter 7) was developed using this case study and cycle counting at locations with no high frequency data was estimated by employing the energy dissipation model, equation 7.7, developed in section 7.5.5. Further details on physical and hydraulic characteristics of this case study, and single-variate analysis of historical pipe failure records in this area are presented in chapter 4.

Case study 2 (CS2) This case study covers the parts of the network for which collected high frequency pressure data was used to estimate CPIS components (details of the sampling programme and estimation of CPIS components are explained in chapters 6 and 7, respectively). The total pipe length in this case study is 1397.43 km, with 6774 pipe failure records.

Case study 3 (CS3) Entire network of the Company is considered as CS3, amounting to approximately 18093.95 km length of pipe, and 35729 number of pipe failure records. Further details on characteristics of the entire company's network is presented in chapter 4.

As the focus of this study is on metallic pipes, only metallic pipe segments and pipe failure records on metallic pipes are considered. Table 8.1 summarises the details of the case studies used. Sample sizes in these case studies are

Table 8.1.: Characteristics of the case studies used for model development

Case study	Number of pipe segments	Total pipe length (km)	Number of pipe failure
CS1	3446	215.04	252
CS2	18502	1397.43	6774
CS3	268285	18093.95	35729

equivalent to the number of pipe segments in each area. Total number of pipe failures are the historical pipe failure records, between year 2003-2013. As logistic regression considers a dichotomous outcome variable (pipe failure or not), pipe samples with n number of failure records are repeated n times in the dataset to represent the impact of predictor variables combination in generating additional outcome variables.

Pipe network is represented in two methods: pipe segment and 50m square-grid. In pipe segment representation, each observation in the model dataset is a pipe segment, of varying length and characteristics. This is the conventional approach. Distribution of pipe segment length is presented in Figure B.1, appendix B.3. Another method is presented here is to use a grid-based representation of the network. For this, a mesh of square-grids is developed to cover the entire pipe network, with each grid inheriting characteristics from the pipe it covers. This approach attempts to normalise the impact from the pipe length by considering a smaller section of the pipe, with less potential variability in characteristics, to be a sample point. Further details on development of grid cell, and impact of grid size are described in appendix B.4.

8.4. Model development

Model datasets are prepared and treated for missing data (detailed in chapter 3), followed by several spatial and single-variate analyses (chapter 4) undertaken to provide initial insight to the relationships between the predictor variables and the outcome variable (pipe failure). Initial list of predictor variables is selected from the initial analysis and also existing literature (conventional predictor variables, listed in chapter 2, and CPIS, detailed in chapter 7). Table 8.2 summarises list of all initial predictor variables considered for model development.

A test of correlation between all predictor variables is undertaken to investigate interaction and potential confounding impact. Model is then fitted as

Table 8.2.: List of predictor variables considered for the models

Variable	Description	Data type
BurstNo	Pipe failure: 0 (no failure) or 1 (failure)	Dichotomous
age	Age	Continuous
diameter	Pipe diameter	Continuous
valveNo	Number of valves on pipe	Continuous
propertyNo	Number of properties supplied by the pipe	Continuous
FP	Soil fracture potential index	Categorical
corrosion	Soil corrosion index	Categorical
material	Pipe material	Categorical
length	Pipe length	Continuous
Pmean	Pipe diurnal mean pressure	Continuous
DP	Dynamic pressure	Continuous

explained in section 8.4.1 and tested based on tests in section 8.2.

There are four models developed, as listed in Table 8.3. Models 1 and 2 are fitted to the dataset from CS1, model 1 using the pipe segment representation approach and model 2, based on grid representation method. The models are developed, firstly, to investigate the impact of various predictor variables and in particular, CPIS, on probability of pipe failure. These models are also used to evaluate effect of two system representation approaches.

Model 3 covers the largest area where CPIS components are estimated using the collected high frequency pressure data, developed to investigate the impact of CPIS components, as well as other predictor variables, on probability of pipe failure. Cycle counting algorithm to calculate cycle amplitude and frequency (as component of DP as predictor variable in the model) is analysed using the collected high frequency pressure.

High frequency pressure data is not available for the entire area used in model 4. Therefore, steady state hydraulic simulations for the entire company network are used to study cyclic pressure changes. The cycle counting algorithm is implemented on the hydraulically modelled data to calculate cycle amplitude and frequencies of pressure fluctuation. It is expected that the steady state pressure changes do not fully represent the dynamic of the network and often underestimate the level of hydraulic instabilities in the network. Comparative analysis of the high resolution pressure data and steady state simulation data in the same areas suggest that in approximately 73% of cases, steady state analysis do not capture the maximum amplitude in pressure profiles. Therefore, the DP calculated from the steady state analysis is a conservative approach of the actual dynamic DP. Any relationship established between the steady state

Table 8.3.: List of models

Model reference	Dataset	Network representation
Model 1	CS1	50m grid
Model 2	CS1	Pipe segment
Model 3	CS2	Pipe segment
Model 4	CS3	Pipe segment

DP and pipe failure probability suggests a similar or stronger relationship in operational hydrodynamic conditions that are exposed to dynamic conditions. Model 4 is developed to study this hypothesis. Moreover, this model is fitted to the largest dataset of the three case studies, providing more certainty in the model outcome.

Below, various steps in development of model 1 are described. Model output, coefficients and various statistics are reviewed and results of several tests are presented. For the other models, similar steps are undertaken in developing and testing the model. The final *preferred* model and the estimated coefficients in each category are presented. This chapter is presenting the model development. Model results and discussions are provided in chapter 9.

8.4.1. Development of model 1

After data preparation, initial data review and analysis, and evaluation of the correlations between all predictor variables, following model development steps are undertaken. Correlation coefficients and plots for this dataset are listed in Appendix B.5, with no strong correlation observed between predictor variables:

Step 1 Development of the null model (model one only considering an intercept with no predictor variables):

```
Null model = GLM (model data, BurstNo ~ 1,
                  family = 'binomial')
```

Adding one predictor variable at a time, starting from the null model, each time, the new model is checked to confirm its significance (detailed in section 8.2.1) and whether an improved model is required.

Step 2 After adding each new predictor variables, one at a time, ANOVA (analysis of variance) test is undertaken:

```
anova(modelA, modelB, test = "Chisq")
```

Table 8.4.: Model coefficient estimates (Model 1)

	estimate	std. error	z-value	Pr($>|z|$)	significance
(Intercept)	-4.497	0.4919	-9.141	$<$ 2e-16	***
age	-0.02182	0.003233	-6.748	1.5e-11	***
diameter	-0.00598	0.001114	-5.364	8.13e-08	***
log(length)	0.9298	0.08695	10.694	$<$ 2e-16	***
materialDI	-1.972	0.2423	-8.137	4.04e-16	***
Pmean	0.01388	0.005871	2.364	0.0181	*
DP	2.15e-05	7.12e-06	3.018	0.00255	**

This shows the change in deviance obtained by adding each of the new terms, deviance change, residual deviance, DOF, and Pr(Chi) for each variable. The sequence of introducing predictor variables to the model does not matter. Models are also compared based on their AIC, Akaike's Information Criterion, explained below, to select the preferred model.

Step 3 The process continues until the preferred model is found.

The preferred model developed for model 1 is:

$$\begin{aligned} ln(OR) = (-4.497) &+ (-0.02182) \times (age) + (-0.005975) \times (diameter) \\ &+ (0.9298) \times (log(length)) + (-1.972) \times (materialDI) \\ &+ (0.01388) \times (Pmean) + (0.00002149) \times (DP) \end{aligned} \tag{8.18}$$

The p, as per section 8.2.1, is large ($\simeq 1$), indicating that there is no evidence of lack of fit.

The estimated coefficients from fitting the data to the model 1 are shown in Table 8.4. In this table, *estimate* in the table is the coefficient estimate, *std. error* is the standard error, *z-value* is the Wald Test statistics with the p. The last column represents the significance of the variable in the model on a 0.05 significance level, each referring to $|z|$ level range as shown in the following table.

Significance	****	***	**	*	.	
$|z|$ level	0	0.001	0.01	0.05	0.1	1

Null deviance and residual deviance are 1803.3 and 1519.2 on 3445 and 3439 DOF, respectively. The distribution of the deviance residuals for model 1 is:

Min	1Q	Median	3Q	Max
-1.2508	-0.42	-0.2592	-0.152	3.4047

Akaike's Information Criterion, AIC, is defined as:

$$-2 \times LL_m + 2m \tag{8.19}$$

where LL is maximised log-likelihood and m is the number of parameters in the model. AIC is the measure of model fit. It penalizes the model for the number of model coefficients. Hence, the preferred model is the one with minimum AIC value. AIC for the proposed model is 1533.2.

Analysis of deviance is undertaken for model coefficients and illustrates the degree of freedom for each coefficient, its deviance, residual DOF and deviance, and its significance. This analysis sequentially compares the smaller model with the next more complex model by adding one variable at each step. Each comparison is undertaken by undertaking a likelihood ratio test. Model output for model 1 is presented in Table 9.1.

8.5. Final models

Following model development and model test steps, list of predictor variables and best model fit are selected. Equation 8.20 shows the general formula for the models 1, 3 and 4.

$$\begin{aligned} ln(OR) = ln(\frac{P(X)}{1-P(X)}) =& \beta_0 + \beta_1.age + \beta_2.diameter + \beta_3.log(length) \\ &+ \beta_4.material + \beta_5.Pmean + \beta_6.DP \end{aligned} \tag{8.20}$$

Model coefficients, β_0 to β_6, are estimated from fitting the model to it corresponding dataset. Fitted models for these models are as following:

Model 1

$$\begin{aligned} ln(OR) = (-4.497) &+ (-0.02182) \times (age) + (-0.005975) \times (diameter) \\ &+ (0.9298) \times (log(length)) + (-1.972) \times (materialDI) \\ &+ (0.01388) \times (Pmean) + (0.00002149) \times (DP) \end{aligned} \tag{8.21}$$

Model 3

$$ln(OR) = (-4.901) + (-0.01189) \times (age) + (-0.006454) \times (diameter) + (1.089) \times (log(length)) + (-1.581) \times (materialDI) + (0.00842) \times (Pmean) + (0.00001828) \times (DP) \quad (8.22)$$

Model 4

$$ln(OR) = (-4.96602) + (-0.00177) \times (age) + (-0.00754) \times (diameter) + (0.9576419) \times (log(length)) + (-1.16394) \times (materialDI) + (0.007368) \times (Pmean) + (0.000687) \times (DP) \quad (8.23)$$

Model 2 is developed using CS2 dataset, with grid cell network representation. The predictor variables for this model, in addition to those considered in above models, include number of valves on the pipe ($valveNo$), but exclude pipe length ($length$) variable.

Model 2

$$ln(OR) = (-3.4855685) + (-0.01302) \times (age) + (-0.00255) \times (diameter) + (0.431064) \times (valveNo) + (-1.85938) \times (materialDI) + (0.017111) \times (Pmean) + (0.742631) \times (DP) \quad (8.24)$$

The results from fitting the models to their corresponding datasets and discussions on these results are presented in chapter 9.

8.6. Summary

In this chapter, logistic regression modelling is used in order to develop four models, fitted to three datasets from the network. Background of the logistic regression modelling, model development and model testing are discussed. In the following chapter, the results from these models are presented and discussed.

9. Model results and discussions

This chapter provides the results and discussions for the statistical models developed in chapter 8.

9.1. Statistical model results discussion

9.1.1. Models statistics

Table 9.1 provides a summary of model statistics from model 1. The large value of p indicates that there is no evidence of lack of fit. Furthermore, the predictor variables reduced the residual deviance to 1519.2 from the deviance of null model at 1803.3, a noticeable reduction, with a loss of 6 DOF. In developing the model, the AIC of 1533.2 is the smallest value achieved when compared to other similar candidate models.

The number of Fisher Scoring iterations indicates that the model converge with 6 iterations. Likelihood ratio test statistic is 284.14 (distributed Chisq) on (6) DOF with Pr(>Chisq)<2.2e-16, indicating that the model fit significantly better than the null model.

McFadden pseudo-R^2 of 0.19042 indicates a good model fit. As suggested by Louviere et al. (2000) and Domencich and McFadden (1975), values of pseudo-R^2 between 0.2 and 0.4 are considered to be indicative of extremely good model fits, equivalent to 0.7-0.9 range obtained for R^2 for a linear function.

Area under the ROC curve for this model is 0.791853. Considering the value of 0.7 as evaluation threshold, this indicates a strong model performance.

Model statistics for model 2 is presented in Tables 9.4. Model 2 is fitted to the dataset from the same case study (CS1) as model 1. The CS1 network in model 2 is represented using grid-based representation, while in model 1, pipe segments are used for network representation. This has resulted in larger number of sample points in model 2. Furthermore, as grid-based representation intends to normalise the impact of pipe length predictor variable *length* is not considered in the model.

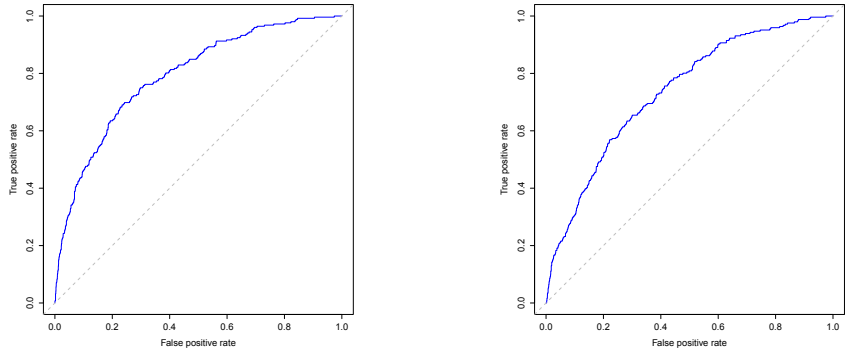
Model 2 has a McFadden pseudo-R^2 of 0.064384, suggesting a poor fit, although the AUROC curve of 0.736566 in this model illustrates a strong model performance. The inferior performance of this model compared to model

Table 9.1.: Model statistics (Model 1)

Model statistic	Output
Dataset	CS1
p (residual's deviance)	$\simeq 1$
Null deviance	1803.3 on 3445 DOF
Residual deviance	1519.2 on 3439 DOF
AIC	1533.2
Number of Fisher Scoring iterations	6
Log Likelihood	-759.6
Likelihood ratio test against null model	284.14 (distributed Chisq) on (6) DOF with Pr(>Chisq)<2.2e-16
McFadden pseudo-R^2	0.19042
AUROC	0.791853

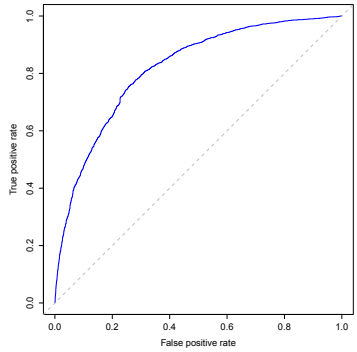
1 is due to the system representation, resulting in larger sample datasets, while the number of success outcome variables (pipe failure) remain unchanged. This leads to significant variation between number of zeros and ones in the dataset, impacting models prediction ability and performance. Development of this model was to investigate the impact of network representation in the model. Based on the results and analysis, the segment representation of the network was decided to be a more reliable approach. This is the representation used for developing models for larger networks (CS2 and CS3).

Tables 9.7 and 9.10 show the model statistics for models 3 and 4. Model 4 is fitted to dataset CS3, that is significantly larger than CS2 dataset used for model 3. Both models show a large residual deviance p (close to 1) and have a relatively high AUROC curve value, 0.816308 and 0.779944, for model 3 and 4, respectively. McFadden pseudo-R^2 for model 4 is 0.1562, a relatively low value indicating a poor fit, while model 3 has a high McFadden pseudo-R^2, value of 0.34998, indicating a very strong model fit. Figure 9.1 illustrates ROC curves for model 1 to 4. Models with high discrimination ability will have low *false positive rate* and high *true positive rate* simultaneously, leading to an ROC curve close to the top left corner of the plot. An ROC curve that is the 45 degree diagonal line suggests that the model has no discrimination ability.

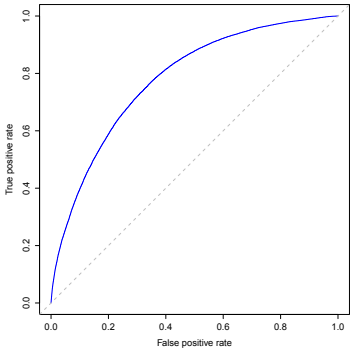


(a) Model 1

(b) Model 2



(c) Model 3



(d) Model 4

Figure 9.1.: ROC curves for the models

9.1.2. Estimated coefficients and the odds ratios

The estimated regression coefficients from fitting the CS1 data to the model 1 is presented in Table 9.2 (repeated from section 8.4.1). The table also shows the standard error for each estimate, its $z - value$ and p, the variable significance on a 0.05 significance level. All predictor variables (except Pmean) are much smaller than 0.05, indicating their significance in rejecting the null hypothesis. For Pmean, the value is larger, but still well below the cut-off point, and considered significant.

Each estimated coefficient in Table 9.2 represents the expected change in log odds, $ln(odds)$, of pipe failure, for one unit increase in the corresponding predictor variables, holding all other predictor variables constant at certain value.

The negative signs for *age* and *diameter* estimates indicate a negative trend between increase in these predictor variables and increase in odds of pipe failure. Considering material is a dichotomous variable (holding values of DI and CI), the negative sign for *materialDI* indicates that there is higher odds of pipe failure for CI. Positive values for the estimates in *Pmean* and *DP* show the expected increases in log odds of pipe failure (0.01388 and 2.15e-05

Table 9.2.: Model coefficient estimates (Model 1)

	estimate	std. error	z-value	Pr($>\|z\|$)	significance
(Intercept)	-4.497	0.4919	-9.141	$<$ 2e-16	***
age	-0.02182	0.003233	-6.748	1.5e-11	***
diameter	-0.00598	0.001114	-5.364	8.13e-08	***
log(length)	0.9298	0.08695	10.694	$<$ 2e-16	***
materialDI	-1.972	0.2423	-8.137	4.04e-16	***
Pmean	0.01388	0.005871	2.364	0.0181	*
DP	2.15e-05	7.12e-06	3.018	0.00255	**

Table 9.3.: Model's Odd Ratio and the 95% confidence intervals (Model 1)

	OR	2.50%	97.50%
(Intercept)	0.011142	0.004205	0.028984
age	0.978421	0.972119	0.98453
diameter	0.994042	0.991698	0.996072
log(length)	2.533935	2.145284	3.017367
materialDI	0.139186	0.084953	0.220398
Pmean	1.013974	1.002114	1.025449
DP	1.000021	1.000007	1.000035

respectively), as the result of one unit increase in these predictors. For length predictor, every one unit change in *log(length)*, (that is equivalent of a 10-fold increase in the length), results in an expected 0.9298 increase in log odds of pipe failure.

Odds is the ratio of probability of pipe failure (p) to probability that pipe failure does not occur ($1-p$). Odds Ratio, OR, is ratio of odds of success (pipe failure) for one unit increase in the predictor variable, holding other predictor variables constant. OR is calculated by exponentiating the estimates of the predictor variables. Table 9.3 shows the OR for the model predictor, and the 95% confidence intervals.

As shown in this table, odds of pipe failure is increased by a factor of 0.978421, for every unit increase in *age* predictor variable. Similarly, for a one unit increase in pipe *diameter*, the odds of pipe failure is increased by a factor of 0.994042. For pipe *length*, 2.533935 is the OR associated to 10-fold increase in length (or one unit increase in *log(length)*), suggesting that the odds of pipe failure will increase by 2.533935-fold, for every 10-fold increase in pipe length.

Pipe *material* being a dichotomous predictor variable in the model, the OR shown for the *Intercept*, the value of 0.011142, is the odds associated to CI pipe material. The value of 0.139186, OR shown for *materilDI*, is the ratio of

odds for DI to the odds for CI, indicating the odds of pipe failure for DI is 0.139186 of odds of failure in CI, i.e. odds of failure in CI is 7.184 times higher than odds of failure in DI.

For *Pmean* and *DP*, odds of pipe failure in increased by a factor of 1.013974 and 1.000021, respectively, for one unit increase in Peam and DP.

Confidence intervals for none of the predictor variables crosses 1, implying that population effect lies within the ranges.

Model predictor variables and the odd ratios for model 2 are presented in Tables 9.5 and 9.6. In this model, number of valves (*valveNo*) on the pipe is found to be significant. A similar set of predictor estimates (sign and magnitude) and odds ratio are observed in this model, compared to model 1, indicating similar impacts from these predictors on predicted probability of failure. Although, the main differences are inclusion of number of valves predictor and exclusion of pipe length as predictor variable. Effect plots (discussed below) for the predictor variables of this model is presented in Appendix B.6.

Tables 9.8 and 9.9 present the model predictor variables and the odd ratios for model 3 and model 4. The predictor variables show similar sign (direction of impact) compare to those in model 1. Pmean and DP are shown to be more statistically significant (lower $\Pr(>|z|)$) in model 3 and 4, although the estimates are slightly lower in these models. Other estimates in these models (except for age and material) are relatively close as well. However, these estimates represent the impact on the log odds of failure. A more intuitive comparison is the OR. The OR for age, diameter, Pmean and DP are very close in these models. The OR for *log(length)* is higher in model 3 compared to model 1 and 4. The difference in OR of *Intercept* indicate that the predicted odds of pipe failure for CI pipes is higher in model 1 than it is in model 3 and 4. Multiplying OR of *materialDI* to OR on the *Intercept* provides odds of pipe failure for DI pipes, which is larger in model 3.

Table 9.4.: Model statistics (Model 2)

Model statistic	Output
Dataset	CS1
p (residual's deviance)	$\simeq 1$
Null deviance	2538.8 on 15885 DOF
Residual deviance	2375.3 on 15879 DOF
AIC	2391.3
Number of Fisher Scoring iterations	8
Log Likelihood	-1187.7
Likelihood ratio test against null model	163.46 (distributed Chisq) on (6) DOF with Pr(>Chisq)<2.2e-16
McFadden pseudo-R^2	0.064384
AUROC	0.736566

Table 9.5.: Model coefficient estimates (Model 2)

	estimate	std. error	z-value	Pr($>\|z\|$)	significance
(Intercept)	-3.48557	0.310729	-11.217	<2e-16	***
age	-0.01302	0.002931	-4.443	8.88e-06	***
diameter	-0.00255	0.000813	-3.136	0.00171	**
valveNo	0.431064	0.059825	7.205	5.79e-13	***
materialDI	-1.85938	0.247013	-7.527	5.17e-14	***
Pmean	0.017111	0.005279	3.241	0.00119	**
DP	0.742631	0.37777	1.966	0.04932	*

Table 9.6.: Model's Odd Ratio and the 95% confidence intervals (Model 2)

	OR	2.50%	97.50%
(Intercept)	0.030636	0.016624	0.056218
age	0.987061	0.981338	0.992689
diameter	0.997454	0.995771	0.998961
Pmean	1.017258	1.006763	1.027821
DP	2.0101458	1.000095	4.212145
materialDI	0.155768	0.094316	0.249119
valveNo	1.538894	1.366466	1.724567

Table 9.7.: Model statistics (Model 3)

Model statistic	Output
Dataset	CS2
p (residual's deviance)	$\simeq 1$
Null deviance	15006 on 18501 DOF
Residual deviance	11911 on 18495 DOF
AIC	11925
Number of Fisher Scoring iterations	6
Log Likelihood	-5955.7
Likelihood ratio test against null model	3094.7 (distributed Chisq) on (6) DOF with Pr(>Chisq)<2.2e-16
McFadden pseudo-R^2	0.34998
AUROC	0.816308

Table 9.8.: Model coefficient estimates (Model 3)

	estimate	std. error	z-value	Pr($>\|z\|$)	significance
(Intercept)	-4.90	0.134	-36.667	<2e-16	***
age	-0.0119	0.000912	-13.036	<2e-16	***
diameter	-0.00645	0.000419	-15.405	<2e-16	***
log(length)	1.09	0.0265	41.068	<2e-16	***
materialDI	-1.58	0.0849	-18.633	<2e-16	***
Pmean	0.00842	0.00126	6.695	2.16e-11	***
DP	1.83e-05	3.20e-06	5.707	1.15e-08	***

Table 9.9.: Model's Odd Ratio and the 95% confidence intervals (Model 3)

	OR	2.50%	97.50%
(Intercept)	0.00744	0.005718	0.009656
age	0.988176	0.986405	0.98994
diameter	0.993567	0.992735	0.994367
log(length)	2.972104	2.822868	3.132199
materialDI	0.20567	0.173847	0.24249
Pmean	1.008455	1.005962	1.010934
DP	1.000018	1.000012	1.000025

Table 9.10.: Model statistics (Model 4)

Model statistic	Output
Dataset	CS3
p (residual's deviance)	$\simeq 1$
Null deviance	210539 on 268284 DOF
Residual deviance	177652 on 268278 DOF
AIC	177666
Number of Fisher Scoring iterations	6
Log Likelihood	-88826
Likelihood ratio test against null model	32886 (distributed Chisq) on (6) DOF with Pr(>Chisq)<2.2e-16
McFadden pseudo-R^2	0.1562
AUROC	0.779944

Table 9.11.: Model coefficient estimates (Model 4)

	estimate	std. error	z-value	Pr(> |z|)	significance
(Intercept)	-4.96602	0.038239	-129.867	<2e-16	***
age	-0.00177	0.000228	-7.766	8.12e-15	***
diameter	-0.00754	0.000172	-43.953	<2e-16	***
log(length)	0.957642	0.006688	143.179	<2e-16	***
materialDI	-1.16394	0.025351	-45.914	<2e-16	***
Pmean	0.007368	0.000385	19.147	<2e-16	***
DP	0.000687	0.000194	3.551	0.000384	***

Table 9.12.: Model's Odd Ratio and the 95% confidence intervals (Model 4)

	OR	2.50%	97.50%
(Intercept)	0.006971	0.006467	0.007513
age	0.99823	0.997784	0.998677
diameter	0.99249	0.992155	0.992823
log(length)	2.605545	2.571683	2.639998
materialDI	0.312253	0.297059	0.328097
Pmean	1.007395	1.006634	1.008154
DP	1.000687	1.000307	1.001066

9.2. Effect plot

One method to assess the model, numerically and visually, is to use the model with some typical values to make prediction. Various combinations of the predictor variables are used to generate the predicted probability of the model. The result can then be plotted to evaluate the change in predicted probability as varying predictor variables.

The concept of *effect displays* for generalized linear models were introduced by Fox (1987) and later expanded by Fox (2003) and Fox and Andersen (2006). An effect display is generated by allowing the predictors to range over their values, while other predictors in the model are held to *typical* values. A reasonable choice for the typical value is the variable mean (Ford, 2016). The result is a list of predicted probabilities for a multiple of values from each predictor variable in its corresponding range.

This algorithm is applied to the models developed, in order to assess and visualise the effect from each predictor variable on the predicted probability of pipe failure. When developing predicted probabilities for each predictor variables, all other variables are fixed at their *mean* values. For *material*, this is fixed at a value equivalent of ratio of number of DI pipes over CI pipes.

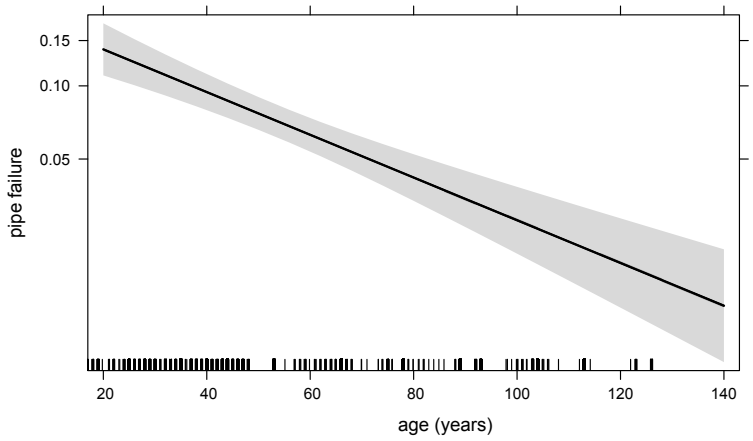
The generated plot illustrates the effect of each variable with its 95% confidence interval band. Furthermore, distribution of the predictor variable values from the dataset is displayed on the plot's x-axis.

Age

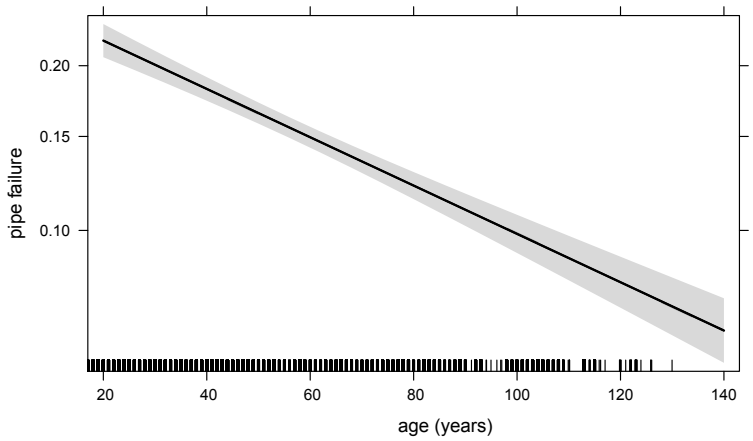
Figure 9.2 shows the effect plots for age predictor in models 1, 3 and 4. This is a somewhat surprising result, as it is contrary to general expectations that older pipes tend to fail more often, although the relationship between age and pipe failure has been found to be complex. Previous studies also showed conflicting results on impact of age on pipe failure rate. Analysis in the study undertaken by Boxall et al. (2007) showed a decrease in CI pipe failure probability with age, for period up to 100 years, although such trend is only expected for a relatively short period of 'burn-in' in bathtub curve, illustrated in Figure 2.2 in chapter 2.

Confidence intervals in model 3 and 4 are smaller than model 1, due to the higher sample size in these two models. Rate of change in predicted pipe failure is considerably lower in model 4, compared to model 1 and 3. Again, this is likely due to the size of dataset in model 4, a considerably larger dataset.

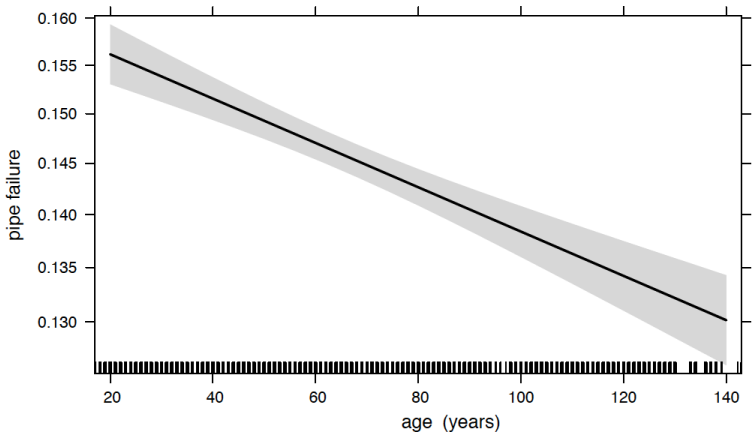
In order to further investigate the effect of pipe age, a new model was fitted to the dataset used with CS2, replacing the age with 10-year age groups. Figure



(a) Model 1



(b) Model 3



(c) Model 4

Figure 9.2.: Effect plot for age

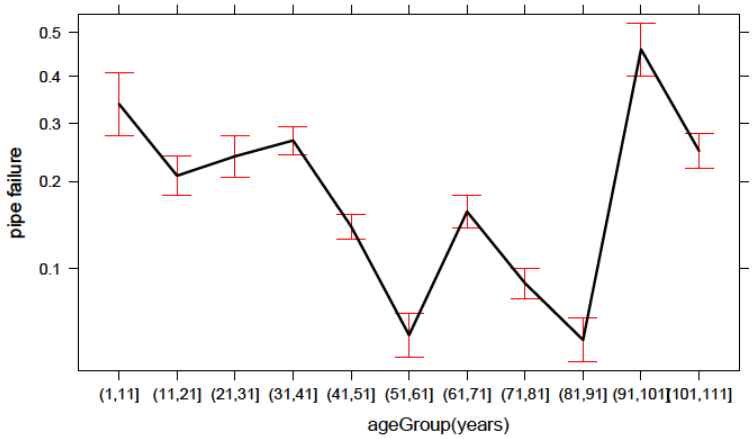


Figure 9.3.: Age group effect plot

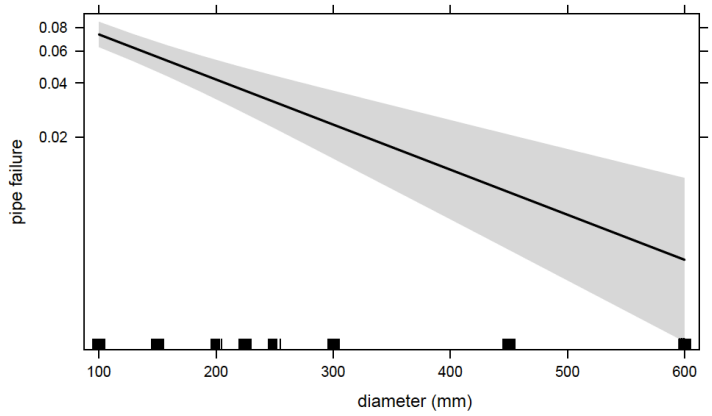
9.3 illustrates the effect on pipe failure probability and confidence interval for various age groups. The figure suggests a sharp decline in expected probability of failure from age group of (31,41] to particularity low probability for pipe age of (51,61] and a similarly low failure probability for age group (81,91], compared to the other age groups. This, again, indicates the complexity of impact from the age on pipe failure. This complexity of relationship is potentially due to the age acting as a surrogate factor for other predictor variables not considered in the model, including design and manufacturing techniques, construction practices and quality of installations, strength and quality of the material, etc.

More research on this topic needs to be undertaken before the association between age and pipe failure is more clearly understood.

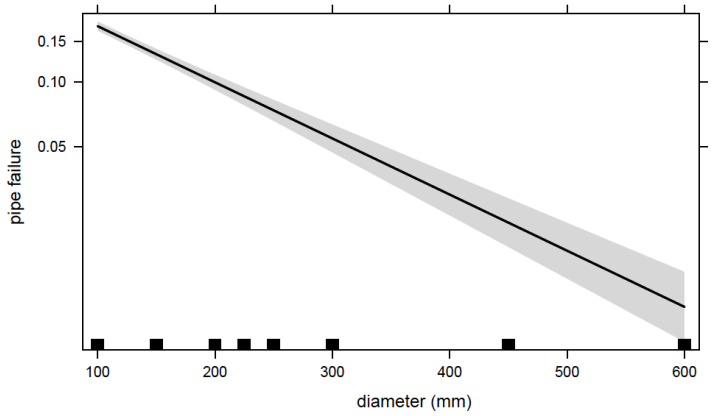
Pipe diameter

The effect plots for pipe diameter in models 1, 3 and 4 are shown in Figure 9.4. All figures illustrate reduction in predicted probability as pipe diameter increases. This is in agreement with the previous literature in this area, e.g. Rajani and Kleiner (2001) and Atkinson et al. (2002). Larger pipes have greater wall thickness, resulting in better performance against deteriorating mechanisms such as corrosion, and structural rigidity due to the larger cross section, leading to enhanced performance in resisting loading and bending stresses.

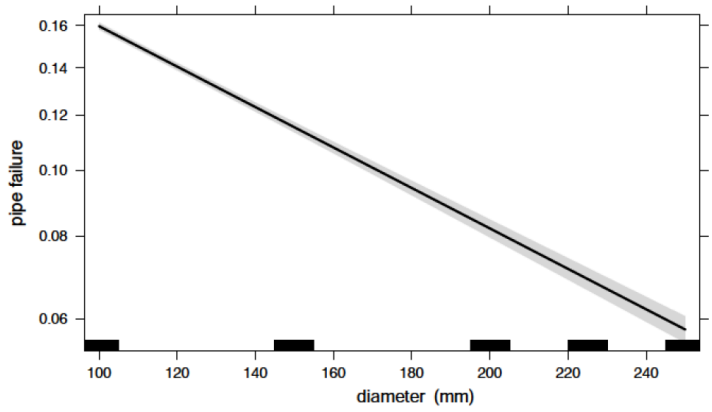
The confidence interval bands in these figure increase with the increase in pipe diameter, owning to there being less sample records in the datasets for larger diameter pipes.



(a) Model 1



(b) Model 3



(c) Model 4

Figure 9.4.: Effect plot for diameter

Pipe length

Figure 9.5 shows the effect plots for pipe length on expected probability of pipe failure. As pipe length increases, there is an increase in the expected probability of failure. The findings of the current study are consistent with those of previous literature, such as Boxall et al. (2007).

The relationship between pipe length and expected probability of failure is found to be curve, rate of increase in failure probability declining with the pipe length rise. Boxall et al. (2007) found a similar relationship in their study.

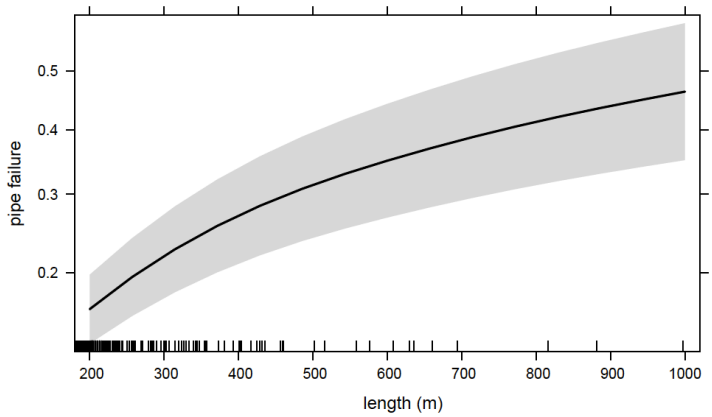
The number of data points lowers as the pipe length increases, the confidence band widens accounting for additional uncertainty in the model prediction.

Pipe material

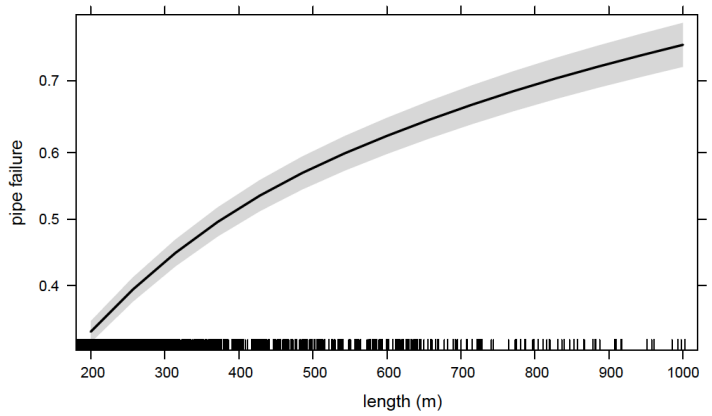
As illustrated in Figure 9.6, CI pipes have higher expected probability of failure than DI pipes. For model 1, these is a value of $\simeq$0.11 for CI versus $\simeq$0.015 for DI. These are $\simeq$0.2 for CI and $\simeq$0.05 for DI in model 3, and $\simeq$0.17 for CI and $\simeq$0.06 for DI in model 4. These are also shown in the ratio of odds of failure for CI against the odds of failure for DI, as described in Tables 9.3, 9.9 and 9.12.

The 95% error bars, indicating the uncertainty in the estimates, shows no overlap between the estimates, suggesting that the difference between the CI and DI is not due to random chance. As size of dataset increases, the error bands shrinks.

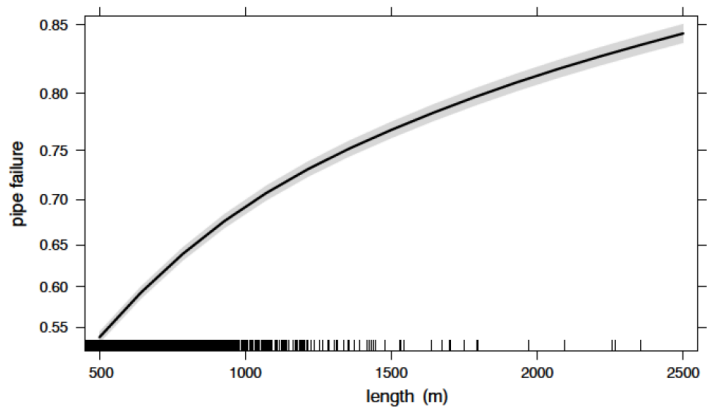
The findings are consistent with other research which found CI to have higher probability of failure compared to DI, for instance Paradkar Brahmanand (2012) and Makar et al. (2001). The brittle nature of CI, combined with degradation due to higher in-service age, poor manufacturing techniques and installation methods are likely to be important contributing factors to this behaviour. Specifically lower temperature (below $5\,^{\circ}C$), increase the CI embrittlement contributing to their brittle fracture and failure. Furthermore, impact from corrosion can exacerbate the deterioration and ultimate failure.



(a) Model 1

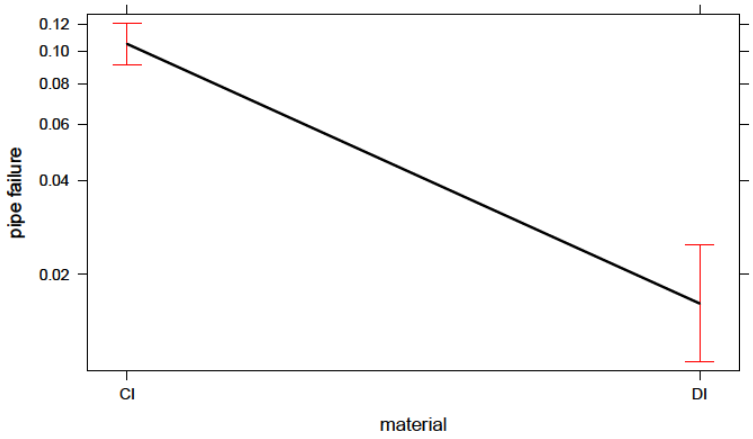


(b) Model 3

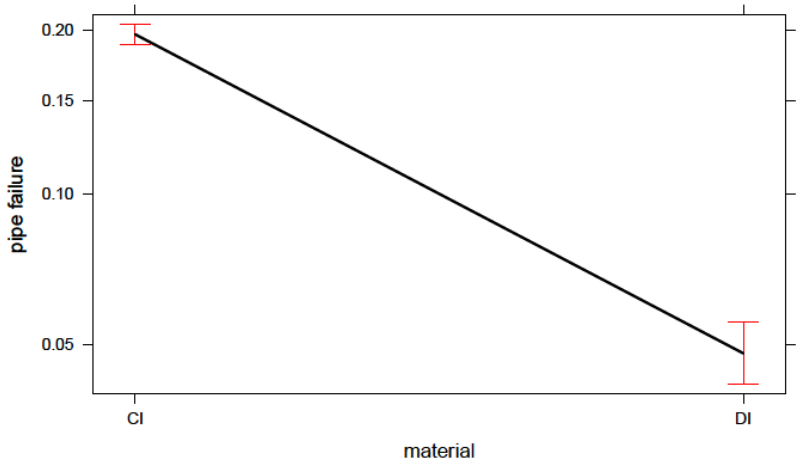


(c) Model 4

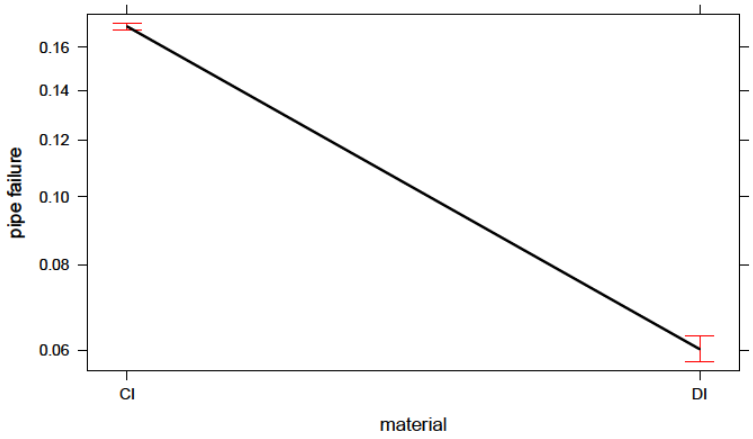
Figure 9.5.: Effect plot for length



(a) Model 1



(b) Model 3

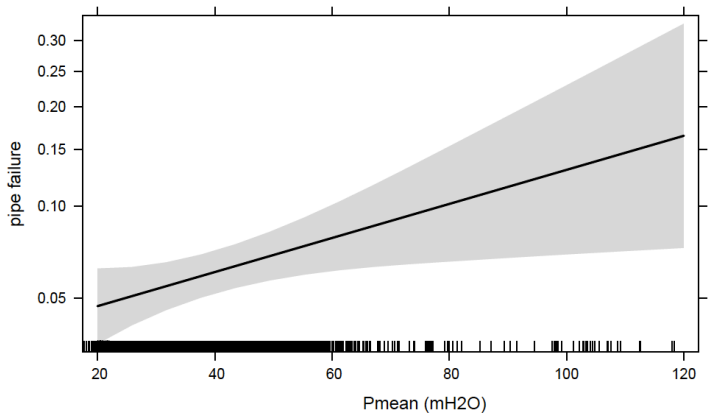


(c) Model 4

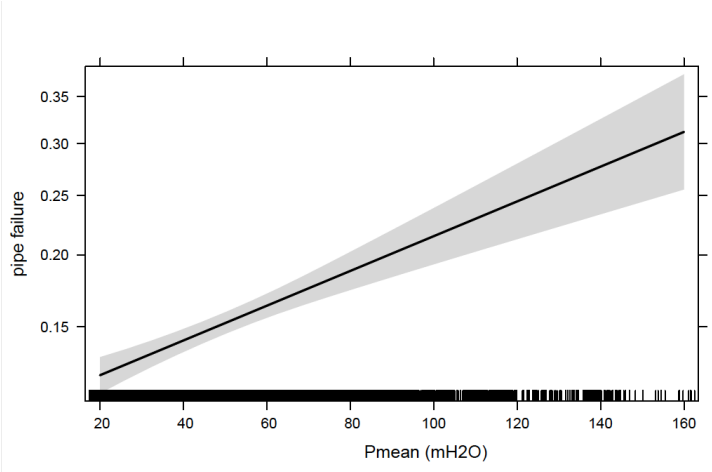
Figure 9.6.: Effect plot for material

Pipe mean pressure

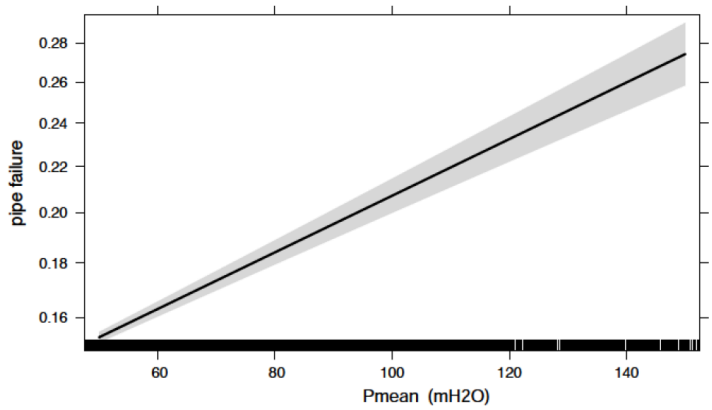
Figure 9.7a shows the increase in expected probability of failure as Pmean increases. The positive trend is suggested in previous studies, such as Lambert (2001) and Thornton and Lambert (2005). Following initial findings in such relationship, most water utilities, including the networks studied here, have extensively implemented pressure management and pressure reduction measures over the last few years. There seems to be additional gains through reduction in expected probability of failure to be had from further reducing the mean pressure. The confidence band widens at pmean extreme values, specifically as it increases, indicating reduced confidence in the prediction in these areas. This is because the majority of pipes have a mean operating pressure of around 40-60mH_2O, and there are considerably less number of pipes with mean pressure outside this range (specifically above 80 mH_2O).



(a) Model 1



(b) Model 3



(c) Model 4

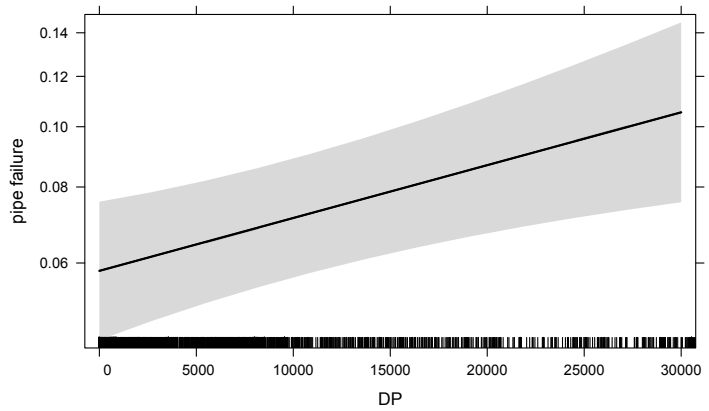
Figure 9.7.: Effect plot for Pmean

DP

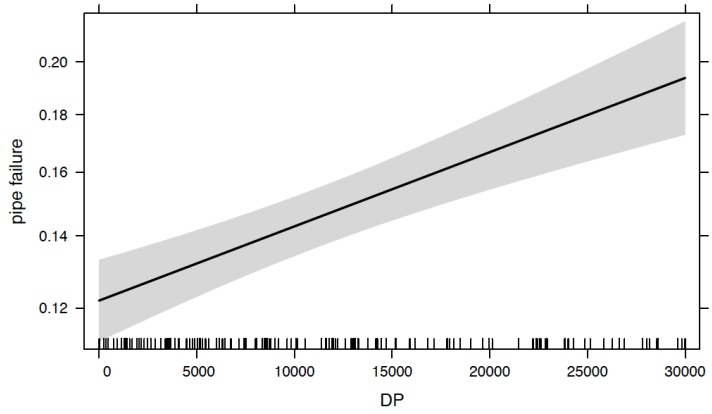
Illustrated in Figure 9.8a, there is a positive association between DP and predicted probability of pipe failure, according to model 1. This is indicating that an increase in DP results in increased predicted probability of failure. Predicted probability of failure escalates from just under 0.06 to above 0.1, with the change of DP from 0 to 30000. This increase is also illustrated in the value of OR (larger than 1), in Table 9.3 for this model.

A similar trend is observed in model 3, Figure 9.8b. A noticeable increase in predicted probability of pipe failure, from approximately 0.12 to just under 0.2 as the result of increasing DP, from 0 to 30000. Widening in the confidence band at the extreme (both minimum and maximum) values of DP is due to smaller data samples in these regions.

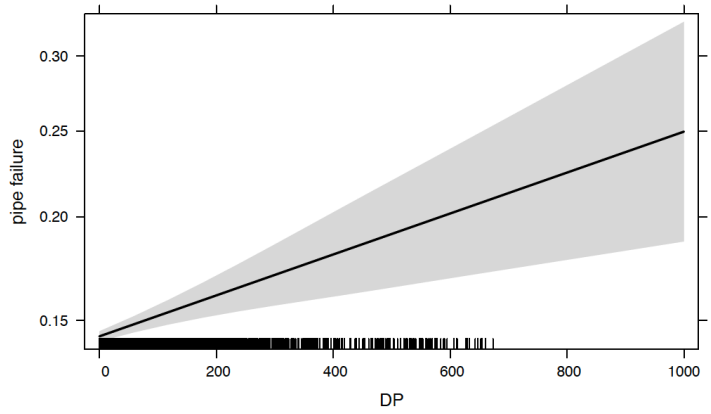
DP used in model 4 is developed from cycle counting of the steady state low resolution hydraulic model output. As such, the distribution of DP values, shown on x-axis in Figure 9.8c, covers a smaller range. A similar observation on association of DP and predicted probability of failure is made in this model too. Confidence band is noticeably larger at higher ends of DP values, due to significant reduction in sample points in this region, i.e. there are limited number of cases where steady state data captures large numbers of high-amplitude pressure cycles. Such data tends to underestimates the level of actual pipe pressure fluctuations, as quasi-steady and unsteady events are not fully captured. Therefore, the calculated DP is often a conservative estimate and the actual figure is expected to be (much) larger. It can, therefore, be safely assumed that higher stresses will only accelerate pipes's degradation and ultimate failure.



(a) Model 1



(b) Model 3



(c) Model 4

Figure 9.8.: Effect plot for DP

9.3. Summary

As associations between DP, mean pressure and predicted probability of failure are established. Considering CPIS in deterioration modelling will clearly enhance the model's ability to analyse the causality and predict risk of future pipe failures across the network. Exact combination of DP and mean pressure to achieve CPIS is obtained from the statistical model coefficients, that is unique to each network and dataset.

The effect of the estimated CPIS and any variations (increase or mitigation to CPIS value as the result of network operation changes) on predicted pipe failure can be analysed using the OR and the effect plot from the model.

10. Cost-benefit framework for calm networks

Steady and unsteady-state pressure variations in water supply networks contribute to degradation and ultimately failure of water pipes. Therefore, systematic evaluation of pipe failure costs and potential benefits from intervention measures intended to mitigate the adverse impact, remains an important aspect of decision making for practitioners across the water industry.

In order to mitigate adverse impacts of hydraulic pressure variations in water networks, utilities need to make various investments to reduce and where possible eliminate such instabilities. This requires the development of a detailed understanding of the supply system, followed by improvements to assets responsible for generation of these conditions. Business justification of intervention measures and activities to eliminate such adverse impacts remain an important factor in undertaking these activities.

The objective of this chapter is to present a methodology for a cost benefit analysis that enables water utilities to evaluate potential outcomes from reducing steady and unsteady-state pressure variations in their networks. This would include the selection of the appropriate monitoring and control technologies, and intervention activities to achieve hydraulically calm networks.

In the chapter, firstly, a review of a number of existing practices, currently used across the industry to evaluate business benefit of the solution schemes is presented.

In addition, the review includes the factors contributing to the total cost of pipe failures (including both direct and indirect costs) and potential cost savings that can be made by reducing such failures. Associated costs of intervention activities to mitigate hydraulic instabilities are also considered.

A novel methodology for expressing operational benefits of future monitoring and control systems as probability distributions to reflect the uncertainty of the estimates is explored with a case study of a large water utility in the UK, reviewing costs and benefits of pipe failures and potential intervention measures to mitigate risk of pipe failures. Net Present Value (NPV) cost benefit approach is employed in combination with Monte Carlo (MC) simulation in

order to provide a methodology to evaluate business justification for cost of intervention activities against benefits from reduction in pipe failures.

10.1. Background

Pipe breaks are one of the three critical metrics used to measure the degree of system optimisation, the other two being pressure management (i.e. AZNP pressure reduction, for measuring hydraulic integrity) and chlorine residual (measuring integrity of water quality) (Water Research Foundation, 2012). Challenges of lifetime cost calculations include predicting future breaks of pipes as well as estimating the range and variations in the associated costs in terms of location, consequence of failure and time.

The whole life cost for a water network include capital costs (design, construction, equipment), operation costs and maintenance costs (including disposal). These costs include internal resources and management, risk allowance, costs for network flexibility, refurbishment costs, sustainability and health and safety costs (Ambrose et al., 2008). Factors to be considered in a whole life cost may include: installation cost, repair cost, leakage costs (background leakage and loss from burst pipes), replacement cost, pumping cost and hydraulic inefficiency, treatment cost, corrosion cost and various direct and societal costs associated to pipe failure. The end of economic life for a pipe is the point at which present value of the future repair costs exceeds the current replacement cost, considering rate of inflation and price changes.

A study by American Water Works Association (2001) highlights a significant need for investment in water assets in the US in the next three decades. While the study is focused on the US network, the findings are applicable to a variety of other systems. The oldest cast iron pipes, dating back to 1800s, have an average life expectancy of around 120 years, while, due to material and manufacturing changes, pipe laid in 1920s would have an average life expectancy of around 100 years, and post World War II pipes are expected to last about 75 years. As such, many of these assets will reach their end of life in the next few decades. The study estimates that by 2030, an average utility in the US would spend three and half times as much, compared to base year of 2001, on pipe replacement (due to wear-out) and three times as much on repairs as pipes get older and prone to break. In another study (American Water Works Association, 2012), it is highlighted that, in the US alone, more than a million miles of pipes are nearing the end of their usual life, requiring to be replaced. The combined costs of this replacement and projected expansion could reach more than $1 trillion over the next two decades, indicating the

global extent of the issue.

Kirmeyer et al. (1994) list the criteria for pipe replacement in the US, as reported by percent of water utilities as a key criterion, to include: number of leaks and breaks (75%), pipe age (45%), low flow (40%), condition or type of material (30%), required size changes (30%), water quality (15%), soil condition (15%), location (10%), street construction work (10%), elimination of dead-ends (5%), and amount of damage by leaks and breaks (5%).

10.2. Costs of pipe failure

There are various costs associated with a burst event. The total cost is combination of *direct* (tangible or internal) costs and *societal* (intangible or external) costs. *Direct* costs are cost directly paid by the water utilities, or their insurance carrier. These might include internal labour, asset and equipment purchase, road and landscape repair, cost of water lost, damage cost as result of flooding, claims paid including legal fees, etc. *Societal* costs are cost either paid by others as actual expenditure and/or in terms of lost wages, interruption to business, lost productivity, etc. Examples of this might include: traffic impact (including people's time in traffic delays), customers supply interruption, interruption cost to dry/wet industry and critical customers, risk of illness and injury, cost of other emergency services (police, fire services), etc. Societal costs could be either hard or soft (Gaewski and Blaha, 2007). Hard costs are paid by the water utilities (directly or indirectly) or sometimes paid by others (e.g. cost of emergency services). Soft costs, such as traffic delays, reduced firefighting capabilities, etc. may never be paid by water utilities, yet it is cost to society. Other environmental costs might be hard or soft, depending on the nature of environmental damage.

Not every pipe burst event has all these costs, and many of the costs listed above (specifically societal costs) might only be applicable to failure of large diameter pipes. In addition, depending on the circumstances of each burst, one cost category might be negligible compared to other categories.

Various attempts have been made to develop tools to incorporate various costs to calculate a total cost of failure. Cromwell (2002) developed a costing model, Grand Central Model, that provides a detailed list of factors to be considered when developing a total cost of failure.

10.2.1. A case study - United States

Gaewski and Blaha (2007) studied 30 large diameter (larger than 20 inches)

pipe failures in the US, primarily CI and pre-stressed concrete cylinder pipe, with failure mode of longitudinal split or broken cylinder. The study found geometric mean of cost of failures for these cases to be approximately $500,000 per failure (ranging from $6,000 to $8,500,000). The total cost of failure for all 30 cases was around $52m with direct and societal costs of $25m and $27m, respectively. For direct costs, more than half of the cost (52%) was for property damage (including minor loses such as water lost, forensic studies, etc.). The outside services (repair contractors, etc.) were the next highest group with 33% of cost followed by internal water utility labour at around 10%, and purchase of material at 5%.

For societal costs, around 57% of the costs was property (building, car, etc.) damages paid by others (e.g. home-owners). Traffic disruption during failure and repair was the next highest group with 27%. This was followed by lost productivity (11%), damage to other utilities (2%), damage to public transportation and reduced firefighting capability at 1% each. The study found burst location, amount of water lost, containment time, and pipe diameter to be good indicators of failure costs, with the first two being the most important indicators.

Gaewski and Blaha (2007) estimated 500 large diameter breaks per year across the US. This, considering an average cost of failure of $0.5m, results in approximately $250m total cost of large diameter pipe failure. Grigg (2007) estimated the cost associated with bursts (most being small diameter mains) in the US to be approximately $10,000 total cost with around $5,000 direct cost. According to this study, in the US, the average annual break rate for all material is estimated to be around 0.25 breaks per mile per year, that is 0.15 breaks per kilometre per year. This is slightly higher than failure rate of this study area (between 0.13-0.14 pipe failures per kilometre per year). Annual number of mains failure for the US are between 250,000 and 300,000 per year, resulting an estimated annual cost of pipe failures to be around $3 billion. Annual flood losses, considered as the most costly natural disaster, are estimated at more than $2 billion of property damage in the US.

10.2.2. Cost-benefit framework

When deciding between proactive and reactive approaches in asset maintenance, post failure reactive costs are of significant importance. Depending on the total cost of failure, main renewal decisions are made, to balance between cost of renewal (decreasing by time compared to present value) and risk of failure (increasing by time due to pipe deterioration). For large (transmission) mains,

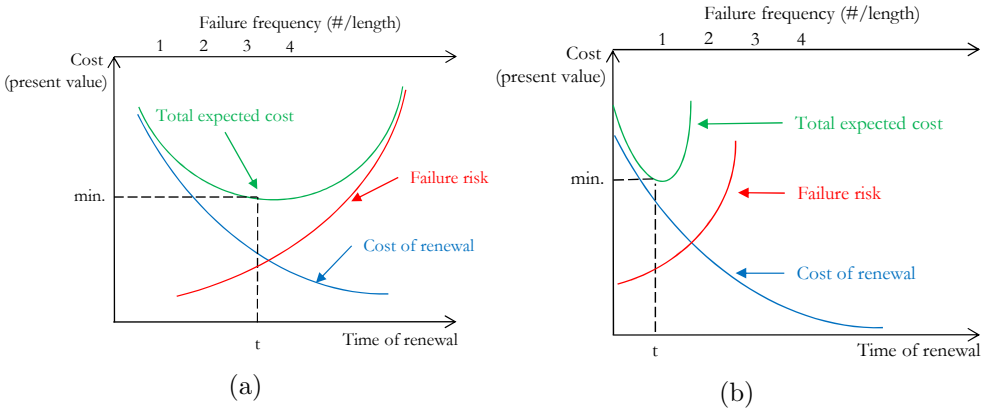


Figure 10.1.: Renewal decision for (a) low cost and (b) high cost failures- adopted from Rajani and Kleiner (2004)

cost of failure, sometimes even a single failure, might be enough to justify renewal costs. Although, the considerable replacement costs of transmission pipes often undermines the risk consideration, Figure 10.1a, while smaller mains might experience a number of failures before economical justifications can be made for their renewal, Figure 10.1b. These figures do not necessarily suggest a linear relationship between the failure frequency and time of renewal.

Water supply networks are exposed to stresses imposed by hydraulic instabilities, contributing to pipe deterioration and ultimately its failure. It is not, however, practical nor economical to mitigate or eliminate all such conditions. Any activities to prevent and/or mitigate adverse impacts of these conditions will incur numerous costs, which should be evaluated against potential benefits from reduction in pipe failures.

Direct and indirect costs associated to any pipe failure and also the range of possible intervention activities and their corresponding cost depend on numerous factors, such as location, pipe characteristics, type of failure, response time, etc. While actual cost of failure depends on numerous factors specific to the case in hand, a cost-benefit framework containing a list of potential adverse impacts and associated costs can be helpful in justifying a business case for proactive measures.

10.3. Methods of cost benefit evaluation

There are a number of methods used across water and other industries to investigate economic justification of a new approach, solution or methodology to tackle an operational challenge. Amongst the most popular ones are Cost-Benefit Analysis (CBA), a Cost-Effectiveness Analysis (CEA), or a multiple-

criteria analysis (MCA). The most popular, CBA, is a systematic way to determine the benefits compared to the costs of undergoing a project in a business (Newell, 2016). Another common approach is the internal rate of return (IRR), which includes a discount factor in the calculation and essentially represents the yield from the investment, however, the IRR ignores the concept of wealth generation, such that it ignores the scale of investment, often not accounting fully for the amount of risk involved (McLaney and Atrill, 2015). One of the most common ratios employed by companies is the Return on Investment (ROI) as it is both simple and versatile. This is the ratio of gain from investment minus cost of investment by cost of investment. The ROI is a measure of the profitability of the investment, but ignores efficiency aspects that can be useful, such as the timing of the return and discount factors. Accounting Rate of Return (ARR) is another common ratio (calculated as average annual operating profit over average investment to earn profit). This is a simple and quick approach, however, it does not consider the time of the return on the investment, and target ARR values are subjective (McLaney and Atrill, 2015). Considering these, benefit cost ratio is considered to study the economical arguments of reducing hydraulic instabilities in the water networks.

10.3.1. Benefit cost ratio

As presented in chapter 9, a positive correlation is observed between CPIS (cumulative pressure-induced stress) from the dynamic behaviour in the system and probability (and total cost) of pipe failure, Figure 10.2a. Mitigating such stresses requires various 'intervention' activities (with their associated costs). Reduction in CPIS level, as indicatively shown in Figure 10.2b, has a positive correlation with costs of associated intervention activities. Shown in Figure 10.2c, the intersection point on the graph indicating a CPIS level at which the cost of intervention is equivalent to the cost of failure (benefit cost ratio of one), defined as 'pressure variation intervention threshold'. Depending on the required benefit-cost ratio, the corresponding CPIS level and pressure variation threshold can then be identified. By increasing positive interventions (resulting in higher intervention costs), the risk and therefore costs associated with pipe failure will decrease.

Benefit-Cost Ratio, BCR, shown in equation 10.1, is the ratio of benefits from pipe failure mitigation to the total costs of failure intervention and prevention measures.

$$BCR = \frac{\text{Failure Reduction Benefits}}{\text{Intervention Costs}} \tag{10.1}$$

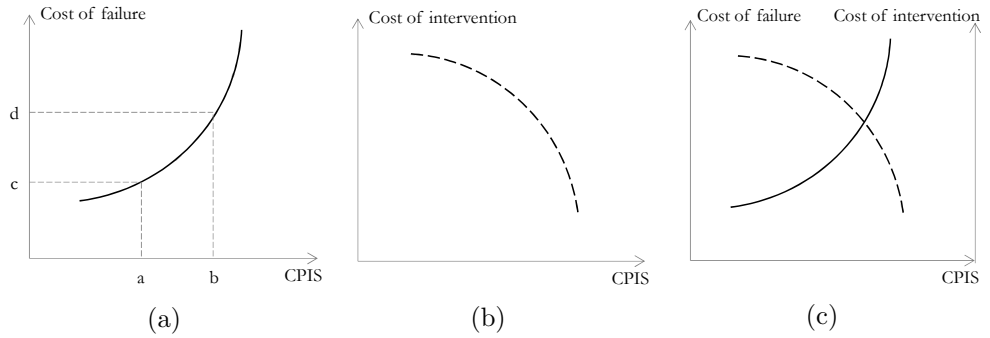


Figure 10.2.: Indicative relationship between: (a) CPIS and cost of pipe burst; (b) CPIS and cost of intervention; (c) costs of pipe burst and intervention for various CPIS

As a theoretical example, in area A, according to the CPIS-burst relationship and depending on the area's characteristics, eliminating CPIS, if at all possible, would result in 10% of burst reduction for small diameter pipes. If the average cost of a small diameter pipe failure in the area is calculated at £2,000 per burst, and the area has experienced 20 bursts in the last 10 years, a reduction of 10% in burst would suggest a saving of 2 bursts in the next 10 years, equivalent of £4,000. If a required BCR of 1.5 is assumed, total of £2,667 is available for intervention activities to mitigate the CPIS, as illustrated indicatively in Figure 10.2a. This is a very conservative estimate as the pipe deterioration will continue during this time period, resulting in increased impact of CPIS in this time period. The intervention measures could potentially prevent additional pipe failures that would have happened if the damaging impact from CPIS continued.

10.3.2. Net Present Value

A common method used for the making business decisions using a CBA concept is the Net Present Value (NPV) formula. The NPV is the summation of all the future cash flows that are discounted at a rate determined by factors of risk in order to determine the present value of a project, equation 10.2 (Arshad, 2012).

$$NPV = \sum_{n=0}^{N_y} \frac{CFn}{(1+i)^n} \tag{10.2}$$

where n is the year from year 0, CF is the cash flow of that specific year, N_y is number of years and i is the discount factor (Arshad, 2012). The approach allows for a discussion on the potential benefits of a new technology or solution

and thus considers the benefits and costs of implementing such technology in a specified time period (Jonkman and Wong, 2001). A positive NPV would then suggest business profitability of the scenarios examined.

In order to optimise a decision, Monte Carlo simulation can be used to provide a range of scenarios. The risk and uncertainty is taken into account through the use of statistical probability distributions correlated with each of the input variables of the NPV calculation. The output of the MC simulation can be assessed to provide the percent of combinations within the determined constraints that give negative NPV values and thus the relative risk associated with investing in the new solution.

10.4. Methodology for cost-benefit framework

The methodology used for this analysis consists of the following steps:

- Estimating benefit of pipe failure reduction and cost of prevention measures (stage 1)
- Distribution and range of cost and benefit estimates (stage 2)
- NPV and MC (stage 3)
- Sensitivity analysis (stage 4)

As shown in Figure 10.3, firstly, the NPV calculation is completed using the optimistic and pessimistic costs, benefits, and growth rates as the inputs to equation 10.2. Employing MC algorithm, a range of random variables are created within the optimistic and pessimistic boundaries, in a probability distribution function appropriate for the respective input variable (Lampka, 2016). The MC simulation also takes into account a sensitivity analysis which will help to determine which costs and benefits specifically are effected the most by changes in the estimations.

10.4.1. Estimating benefits of pipe failure reduction and costs of prevention measures

The main challenge in developing any cost estimation method in the water industry is lack of reliable information and data on factors impacting the total cost. Various studies, such as Cromwell (2002) and Gaewski and Blaha (2007), have studied costs of infrastructure failure and undertaken various cost modelling attempts.

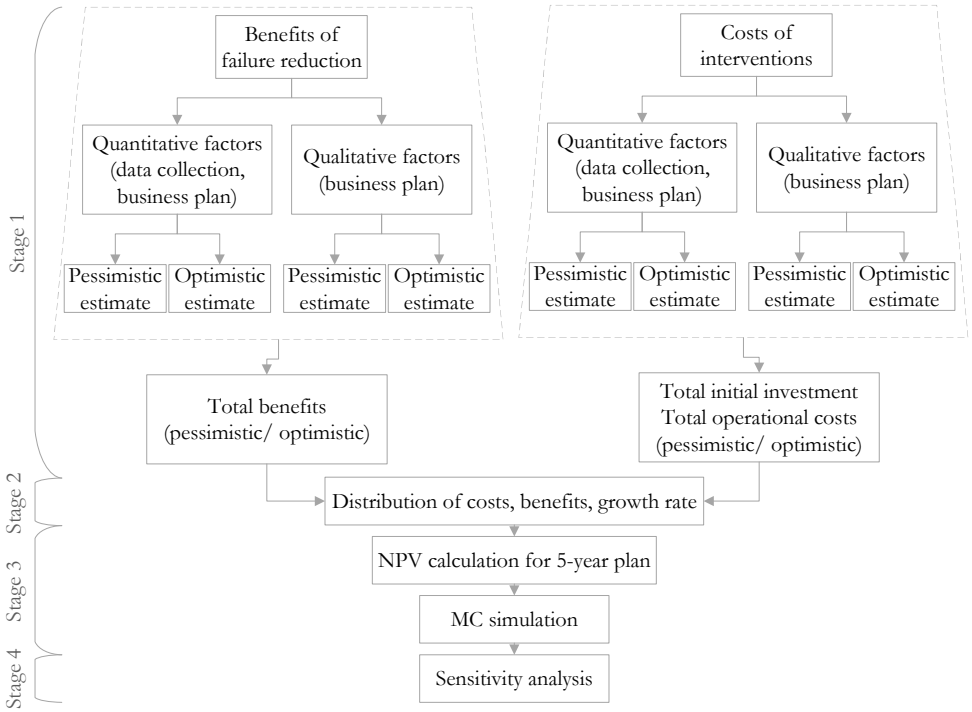


Figure 10.3.: Overview of methodology stages

In this work, a combination of data survey and analysis, questionnaires, utility's business plan and investment strategy, expert interviews and engineering judgement is utilised in order to provides estimate and ranges for such costs, where deterministic figures are not available. Various factors that form the total direct cost of pipe failures are estimated, based on such data. For regulatory cost (penalty) of supply interruptions, pipe failure events are categorised based on pipe diameter size as well as land use (urban/ rural) of where the pipe is located. A set of samples are selected from each cohort to represent the entire pipe burst dataset, and cost of supply interruption is calculated for samples and the entire population (all pipe failure events).

Furthermore, various factors contributing to total costs of intervention activities are reviewed and their costs estimated. Details of these are presented in sections 10.5 and 10.6.

10.4.2. Distribution and range of cost and benefit estimates

The estimated costs and benefits carry a potentially high level of uncertainty. This is firstly due to the natural variations that exists in each case due to its specific characteristics. In addition, many of these estimates rely on engineering judgement and as such are prone to error. Furthermore, in any case, the cost and

benefit figures are *estimates*, hence they carry an inherent level of uncertainty. Therefore, it is necessary to develop a range (lower and higher bands) and an appropriate distribution for each estimate.

Distribution

In this study, a log-normal distribution was chosen for the benefits, while a uniform distribution was chosen for the costs. The benefits use log-normal as it allows for a conservative approach to analysing the uncertainty. A uniform distribution models the same likelihood of occurrence across the range of cost values (Lampka, 2016), as many of the costing values come from the marketplace values, and thus there is a higher confidence that the range of costs will lie primarily within the estimated range at equal probabilities. Geometric distribution was used for the growth rate, as it is generally believed to be an accurate portrayal of reality (Halvarsson, 2013). The geometric distribution puts emphasis on the first years growing more quickly with a gradual exponential decrease in growth rates over time. For the water industry, this seems accurate as when analysing past empirical data, the growth rates of the company over an Asset Management Period (AMP) period seem to have the same trend (Ofwat, 2015). A log-normal distribution was taken for the cash flow and the NPV as the distribution has a lower bound of zero. This is an important characteristic for the NPV as it allows for a conservative approach, such that it will give emphasis on the negative values.

Ranges

In order to develop lower and upper band for values of each factor affecting cost and benefit estimates, a conservative (pessimistic) and an optimistic approximation are considered. In quantitative factors (see section 10.5.), for benefits from burst reduction, the conservative value is usually zero (no saving can be made). Optimistic upper band highlights the maximum realistic percent of the cost that can be mitigated by employing the new approach. This is usually between 2-15%. For the costs of intervention measures, the optimistic is 90% (i.e. delivery of the intervention measure at cheaper rate than anticipated), while the conservative approach is usually 110%, i.e. implementation cost exceeds the forecast. In qualitative factors, where the cost is derived as percentage of relevant expenditure committed by the utility, the lower (pessimistic) band is at zero percent for benefit and 5-15% for intervention costs. Optimistic band is around 2% for benefit and usually zero percent for the intervention cost.

10.4.3. NPV, MC and sensitivity analysis

NPV is undertaken over a 5-year period, equivalent of an AMP period, considering a discount rate. UK treasury recommends a discount rate of 3.5% for schemes shorter than 30 years (Cabinet Office, 2013), however, in this study, a higher discount rate of 10% is used, providing conservative returns, as the study presumes a short period of time. For longer time periods, smaller discount rate should be used.

Using MC simulation, 100,000 different scenarios are simulated in order to estimate risk of non-profitability of the intervention measures. Sensitivity analysis is then undertaken to investigate robustness of the methodology.

10.4.4. Case study data

In developing each step, data and information on assets, pipe failures, costs, etc. are used from a large water utility in the UK. In addition, where available, experts' opinions from practitioners from another UK water utility are utilised to enhance reliability of the estimates.

10.5. Benefit of pipe failure mitigation

The benefit of pipe failure mitigation is from the saving made as a result of reduction in number of pipe burst events. Considering the direct and societal costs of pipe failure, in this study, the cost factors are categorised as the *directly quantifiable* and the *qualitative* costs, as listed in Table 10.1.

The cost estimates for the quantitative factors are derived from the pipe failure direct cost estimates (detailed in Section 10.5.1). These are obtained through data collection, interviews with experts and questionnaire. Quantitative factors are estimated mainly by considering each factor as percentage of the Company's investment plans in the related category, Table 10.2, obtained from the Company's business plan (Severn Trent Water, 2015). For each item, an *importance of impact* (value 1-3) and *occurrence likelihood* (value 1-3) is considered to give weighting to probability and impact of each qualitative item to happen. Values for these items are estimated based on investment plans, importance of impact and occurrence likelihood. For both factor types, a pessimistic (lower band) and an optimistic (upper band) are considered in order to provide a range for each item. Total benefit is then calculated by combining the quantitative and qualitative estimates.

Table 10.1.: Factors impacting total cost of pipe failure

Key benefits	Quantitative	Qualitative
Background leakage reduction	Water loss Leakage maintenance Outcome Delivery Incentive (ODI) benefits Labour costs Emergency projects SIM (Service Incentive Mechanisms) Score benefits	Reputation Environmental benefits Reducing the need for new resources (eg building new reservoirs)
Direct repair cost	Cost of outside services (contractor/ consultant/ laboratory) Bursting valve replacement Pipe replacement material Emergency projects Labour costs Additional bursts (communication pipe bursts) Cost of alternative supply (tankers/ bottled water)	Reputation Social benefit Environmental benefits Less water scarcity Interruption of the service Traffic disruption Other utilities disruption
Other operating costs reduction	Bad debt Energy Insurance/legal Transport Improved customer service /customer complains reduction	Reputation Environmental benefits Market share
Extended asset lifetimes	Capex deferral Optimizing Capex use Maintenance Pipe replacement Labour costs ODI benefits	Reputation Environmental benefits Interruption of the service
Flooding, improved customer service	Flooding SIM score ODI benefits	Reputation Market share Impact on services (gas & electricity) from flooding & repair
ODI supply interruption	Claims from service interruption ODI benefits	Reputation Environmental benefits Market share
Water quality	Optimizing chlorine residual Pipe damage Discolouration	Environmental benefits Public health impact Less water scarcity

Table 10.2.: Investment expenditure from the Company's business plan

Item	£, million
Water quality expenditure	74.4
Demand management expenditure	51.04
Environmental expenditure	97.1
Customer service expenditure	12.8

10.5.1. Failure's direct cost estimation

The quantitative part of benefit from pipe failure reduction comes from the savings made as the result of reduction in direct cost of failure. Direct cost of failure is estimated by using historical data as well as collecting data through questionnaires, experts' interview and engineering judgement. These costs are estimated in the following categories:

- Leakage and active leakage control activities
- Direct repair cost
- Asset life impact
- Flooding and customer service impact (ODI)
- Water discolouration incidents
- Supply interruptions (ODI)

Questionnaire A questionnaire was developed, based on a study undertaken by Gaewski and Blaha (2007), to collect data on the costs associated with recent pipe failure events. The questionnaire was distributed amongst several experts in the Company, and also supply chain and service providers. Reliable data on various elements of pipe failure cost is very limited and a number of questions in the original questionnaire are either removed or combined with other questions, due to limitation of available data. A copy of the final questionnaire is in Appendix C.1. Data gathered from these questionnaires are used in conjunction with other information recorded from other sources (corporate databases, experts' interviews, engineering judgement, etc.) to develop estimation for the cost factors below.

Cost of leakage According to International Water Association studies, as reported by Ambrose et al. (2008), 95% of burst leakage is reported and on average, for a system with nominal operating pressure of $50mH_2O$, the

aggregated leakage losses will be around 12 $m^3/hour$ for 3 days before the leak is repaired. This results in a total loss of $864m^3$. For un-reported bursts (remaining 5%), for a system with nominal operating pressure of $50mH_2O$, with average loss of 6 $m^3/hour$ for 50 days before the repair is undertaken, resulting in total of $7200m^3$ losses.

Using the figures above, and based on the average number of mains repairs in the Company, an estimated cost of water loss as a result of leaking pipes is illustrated in Table 10.3.

Table 10.3.: Annual cost of water lost from leakage

Item	Value
Water loss from reported burst (m^3)	864
Average annual number of bursts	6000
Number of bursts reported (95%)	5700
Water lost from reported bursts (Ml)	4924.8
Number of bursts not-reported (5%)	300
Water loss from not-reported burst (m^3)	7200
Water lost from not-reported bursts (Ml)	2160
Water lost total (Ml)	7084.8
Marginal cost (£/Ml)	£100
Total cost of water lost (£)	£708,480

Active leakage control involve a range of activities to *find and fix* leaks. These include various methods of leak detection and pin-pointing as well as several mains repair costs, asset costs, labour, etc. While the exact cost is highly dependent on various factors specific to each case, an estimation based on historical figures suggests a cost in excess of approximately £330,000 for a reduction of 1 Ml/day of leakage, according to Severn Trent Water Asset Strategy (2015). The Company's leakage target for end of AMP6 is to set to achieve 424 Ml/day, against the current level of 450 Ml/day. This is equivalent to an average reduction of 5 Ml/day reduction for each year in the AMP. The cost of this reduction for each year would add up to £1.65m per year. Marginal cost of water is relatively low at around £100 for 1 Ml of water, while cost per 1 Ml of headroom is approximately £300.

Cost of mains repairs The direct cost of pipe repairs (including labour, material, etc.) is estimated to be on average approximately £1700 and £2700 for small (<250mm) and large (≥250mm) diameter pipes, respectively, according to Severn Trent Water Asset Strategy (2015) and findings from the questionnaire. These costs only consider direct cost bore by the Company for the repair

Table 10.4.: Direct cost of pipe repair

Pipe diameter size	Small (<250mm)	Large (≥250mm)
Average number of bursts	6645	211
Unit direct repair cost (£)	1700	2700
Total annual cost (£million)	11.30	0.57

activities and do not include other indirect and societal costs.

Average number of pipe failure over the period of year 2003 to 2013 for small (<250mm) and large (≥250mm) diameter pipes are 6645 and 211 per annum, respectively. This equals to a total of £11.30m and £5.7m repair costs for small and large diameter pipes, as illustrated in Table 10.4.

Asset life impact By mitigating stresses imposed by hydraulic instabilities, a pipe's deterioration process is delayed and its life extended. Extended pipe life resulting in a reduction in direct costs such as labour and material costs associated with pipe replacement. It is estimated that, on average, 10% of pipes will benefit from extended life, resulting in 10% savings in direct labour cost and pipe replacement material, according to Severn Trent Water Asset Strategy (2015).

Other benefits from pipe asset extension, such as Capex deferral, reduced maintenance and positive impact on various ODI's are estimated as percentage of the Company's investment expenditure as set out in its business plan. These benefits could be considerably greater than other savings such as leakage and direct repair costs.

Cost of flooding and customer service impact It is estimated that approximately 1% of pipe failures large diameter pipes result in a flooding incident causing damage and customer service impact. Intensity and impact of such flooding incidents are greatly varied, depending on characteristics of the pipe failure incidents and the surrounding environment. An average value of £10,000 is considered as total cost of each flooding incident (Severn Trent Water Asset Strategy, 2015).

Positive impact of SIM score and other ODI's as result of mitigations in flooding impacts are considered as percentage of the Company's investment expenditure as set out in its business plan.

Cost of water discolouration events On average, there are approximately 7,500 customer contacts related to water visual appearance, taste, odour and

illnesses in the Company. Pipe failures and hydraulic instabilities can result in dislocation of sedimentation inside the pipe resulting in discolouration. Resulted customer contact and complaints regarding discolouration had numerous impacts on the water utilities, including damage to reputation, regulatory penalties, response activities, etc.

There are, however, several other factors that can contribute to discolouration events. Furthermore, initial analysis of available customer contacts data highlight potential uncertainties. Associating spatially and temporally distributed customer contacts to an individual discolouration event often results in unreliable outcomes. Due to these uncertainties and unavailability of reliable cost estimates of customer contacts, benefits achieved from reduction if discolouration events are not included in the total benefits of pipe failure reduction. This is a conservative assumption, i.e. the benefit will always be greater than these estimates, as any reduction in customer contacts will have a positive impact on the benefit achieved.

Cost of supply interruptions Each supply interruption event, depending on the event duration, attracts various costs associated with response activities, customer complaints, regulatory fines, etc. In many pipe failure events, such costs are the greatest compared to other direct costs discussed above.

In order to estimate total cost supply interruption as result of pipe failure, information and data for one full year has been analysed. These include data on all pipe failure events that result in supply interruption incidents, including duration and impact of each incident, and potential regulatory fine (ODI) as a result if the incident. The database consists of a total of 505 incidents recorded in one year. Each record represents one supply interruption incident which might be the result of multiple pipe failure causing supply interruptions for various number of properties and for different period of interruption time (usually three categories of $\geq$ 3 hours, $\geq$ 6 hours, or $\geq$ 12 hours). Supply interruption from each pipe failure is directly relevant to the pipe diameter size and the land use where the failure occurs, resulting in varying costs.

In order to estimate these costs, four cohorts of pipe failures have been developed, based on pipe diameter size and land use area where pipe is laid. Cohort 1 (C1) represents pipe bursts on small (<250mm) diameter pipes in urban areas, C2 for large ($\geq$250mm) diameter pipes in urban areas, C3 is for small diameter pipes in rural areas and C4 is for large diameter pipes in rural areas, as shown in Table 10.5.

Each cohort consists of a variety of events, with low to high impact incidents. Distribution of property-minutes (P-M: total of number of properties times the

Table 10.5.: Characteristics of pipe cohorts

Cohort ID	Pipe diameter	Land use
C1	<250mm	Urban
C2	≥250mm	Urban
C3	<250mm	Rural
C4	≥250mm	Rural

Table 10.6.: Property-minutes values for each cohort

	C1	C2	C3	C4
Number of incidents	160	17	299	29
Minimum P-M	189	1,850	194	400
Maximum P-M	379,710	1,470,289	573,317	10,004,868
Average P-M	24,077	195,237	33,596	571,782
Median P-M	10,496	36,012	10,500	83,087
Standard deviation P-M	53,200	368,387	63,859	1,855,304

period of time interrupted in minutes) for each cohort is as shown in Table 10.6. Figure 10.4 illustrates distribution of property.minutes interruptions across each cohort.

A sample set from the entire dataset of 505 events is selected in order to gather detailed cost data. Sample size, using equation 10.3, is calculated for confidence interval of 90% and 15% margin of error.

$$SampleSize = (z - score)^2 \times SD \times \frac{1 - SD}{(Marginof Error)^2} \tag{10.3}$$

The values in Table 10.7 are used to calculate the minimum required sample size:

Table 10.7.: Values used in sample size calculation

α	CI	z-score	Margin of error	Standard deviation	Min. sample size
0.1	90%	1.645	15%	0.5	30

Considering proportionate stratification of taking sample from each cohort, the number of samples from each cohort is calculated, using equation 10.4, as show in Table 10.8.

$$n_h = (\frac{N_h}{N})n \tag{10.4}$$

Final sample size is adjusted to account for the lack of (gap in) the available

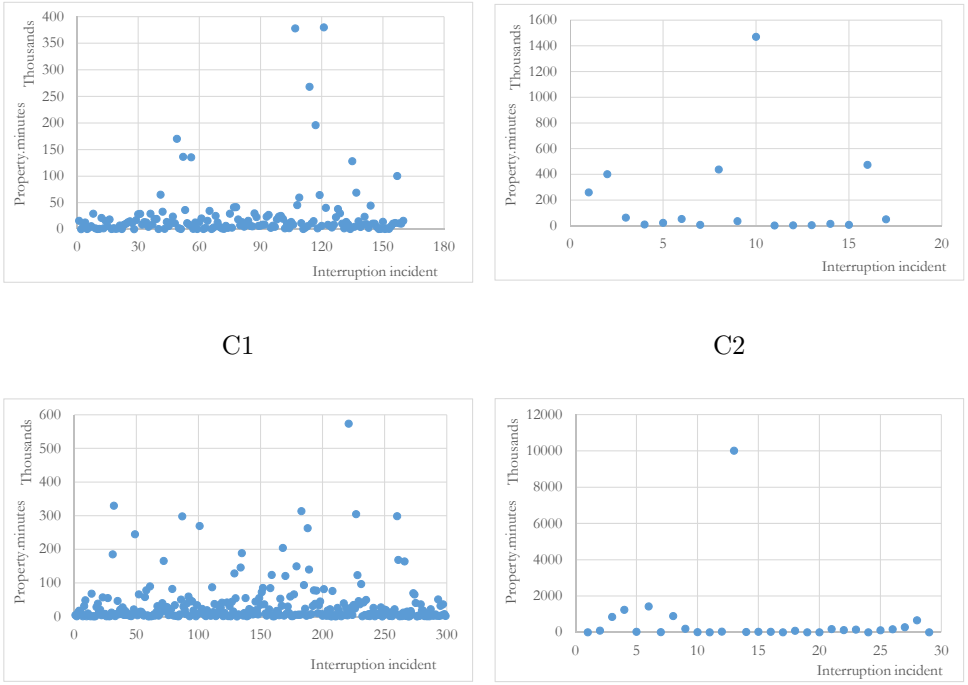


Figure 10.4.: Distribution of property.minute for interruption incidents in each cohort

data for randomly selected cases and to ensure each cohort has a minimum of five samples.

Detailed analyses on cost of failure, specifically costs associated with regulatory penalty (often a significant cost) are undertaken on the sample points. Due to commercial sensitivity of such data, the detailed calculation for these costs are not available, however, the outcome from these analyses are feeding the overall calculation of pipe failure costs.

Each supply interruption event was reviewed and related data was analysed in order to estimate its costs. The total cost is mainly driven by the ODI-related regulatory fines. The cost of each event is proportionate to its contribution to the total property minutes interruption of the Company. Distribution of supply interruption costs for various events in each cohort is shown in Figure 10.5 with average cost in Figure 10.6.

It should be noted that these figures provide estimates for costs of supply interruption incidents. Depending on the targets set by the regulator (Ofwat), the actual costs could potentially vary. Each water utility is required to maintain its supply interruption metric (property.minute) below the threshold

Table 10.8.: Sample size of pipe failure cohorts

Cohort	Number of failures	Min. sample size	Adjusted sample size
C1	160	10	14
C2	17	1	7
C3	299	18	13
C4	29	2	5

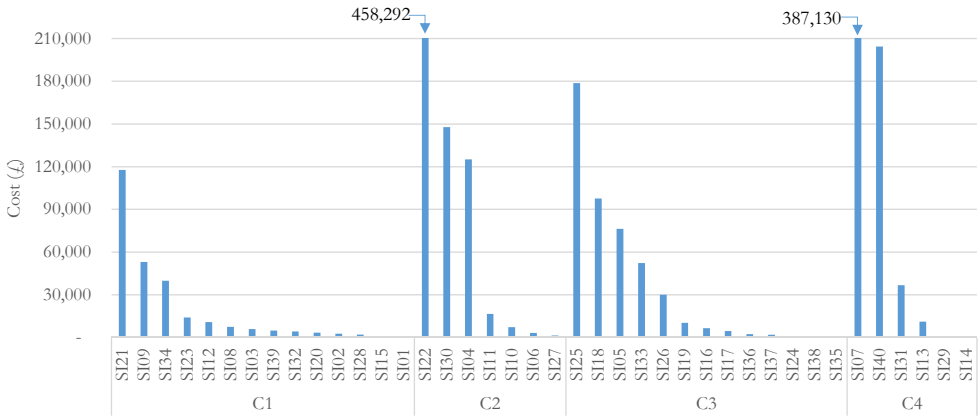


Figure 10.5.: Supply interruption costs of samples in each cohort (labels represent costs outside chart axis range)

set by Ofwat. Any additional interruption above the threshold results in significant penalty. Similarly, achieving an interruption level below the target threshold results in financial rewards for the utility.

Total cost of pipe failure Based on the costs estimation for individual items above, the total cost of failure for various cohorts are estimated. Considering the pessimistic and the optimistic levels, cost range for each factor and a range for the total cost of failure is developed. This forms the benefit of pipe failure reduction, as presented in Table 10.10.

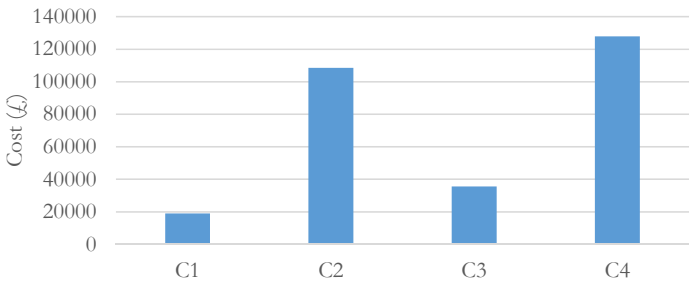


Figure 10.6.: Average supply interruption cost for each cohort per annum

10.6. Cost of intervention activities

Costs of investment in intervention activities include numerous items that can be grouped in the following main categories:

- Equipment installation and IT infrastructure
- Maintenance of new technology
- Unrealised expectations
- Risk of losing customers
- Uncompleted projects

There are several items in each category either be directly *quantitative* or *qualitative*, listed in 10.9. A large proportion of quantitative items consist of mitigating and removing sources of hydraulic instabilities in the network. This includes problem identification (data gathering, system review, etc.), solution development, asset modification and replacement, and operation and maintenance. Main sources of hydraulic instabilities include pumps, automated valves, industrial large consumers, planned and unplanned interruptions, and manual operational activities in the network (valve operation, etc.).

A survey of the water Company and its suppliers is undertaken in order to provide estimates for the quantifiable items. The information also provides a pessimistic and an optimistic range, to account for differences in each case (location, asset, etc.). Where direct figures are not available, percentage of relevant investment from the Company's investment plan is considered as the item cost. Similar approach is undertaken for the qualitative items.

Total costs of intervention activities, separated to *initial* and *operational* investments is calculated by combining the costs of quantitative and qualitative items. These are presented in Table 10.10.

10.7. NPV and MC simulation

Estimated ranges for benefits and costs (initial and operational) as well as growth rate are presented in 10.11. Random values for each parameter are generated (in each model run), based on the distribution relevant to each factor. NPV is then calculated for a 5-year plan (equivalent of an AMP period), using equation 10.2. Assumptions are made for development of the project in the first and second year of AMP period. Full benefit is achieved in the third year, followed by project slowing down in year four and five. These are presented in

Table 10.9.: Key costs of intervention measures

Key Costs	Directly Quantitative	Qualitative
Project installation and development	-Hardware (logger, server, etc.) -Software (IT infrastructure, analytical tools, etc.) -Asset operational regime modification and consultation -Labour -Asset operation assessment -Asset change/modification -Staff training (internal & external)	-Extra safety measures -Advanced planning -Social cost such as traffic -Carbon footprint -Interrupting the service
Maintenance of new technology	-Labour -Replacement costs	-Interrupting the service
Unrealised expectations	-New unexpected operating costs -Less leakage reduction than goal -Training on the system -Other negatives	-Lack of information causing work to fail
Risk of losing customers	-Retail separation effects -Installation costs increase water cost	-Reputation -Market share
Uncompleted projects	-Failed investment -ODI driven penalties	-Reputation from investors

Table 10.10.: Total failure reduction benefits and intervention costs

			£million
Benefits of pipe failure reduction		Optimistic	70.49
		Pessimistic	5.41
Intervention costs	Initial investment	Optimistic	-10.86
		Pessimistic	-33.06
	Operational	Optimistic	-2.46
		Pessimistic	-11.07

Table 10.11.: Range of values used for calculating NPV

	Benefits	Initial costs	Operating costs	Growth rate
Low	5.41	-10.86	-2.46	0.009
High	70.49	-33.06	-11.07	0.003
Random	64.39	-15.42	-4.48	0.0053

Table 10.12.: Calculating present value for each year of 5-year plan

Year	Note	Discount factor	Cash flow	Cumulative cash flow	Present value
0		1	-15.42	-15.42	-15.42
1	First year to achieve 1/3 of benefits	0.91	16.86	1.45	15.33
2	First year to achieve 2/3 of benefits	0.83	38.22	55.09	31.59
3	First year to achieve all benefits	0.75	38.43	76.65	28.87
4	Project slows down by 1/3	0.68	25.5	63.93	17.42
5	Project slows down by 2/3	0.62	8.46	33.96	5.25
	Total				83.04

Table 10.12. Sum of present values for each year of the project plan results in the NPV for this run. This is 83.04 for the run presented here.

Monte Carlo simulation

MC simulations have been carried out, using MATLAB programming software, (MathWorks, 2015), running 100,000 scenarios based on values generated using log-normal distribution for benefits, uniform distribution for costs, geometric distribution for growth. NPV is then calculated for each year of the 5-year study period and the sum of NPVs is used for NPV of each run. Table 10.13 presents the minimum, maximum and average values for cost, benefit and NPV from all runs of MC simulation.

A histogram of NPV from the MC simulation is presented in Figure 10.7. Analysing the frequency of negative NPV shows a total of 19,167 runs resulted

Table 10.13.: Cost, benefit and NPV from MC simulations

	Cost	Benefit	NPV
Minimum	-11.070	14.773	-30.342
Maximum	-2.460	85.472	135.060
Average	-6.763	32.400	9.113

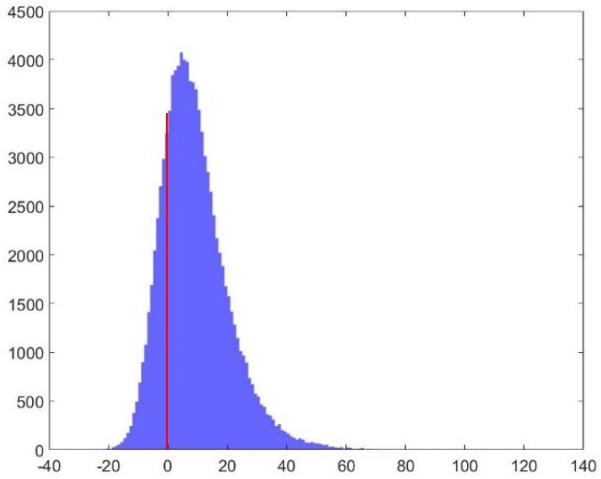


Figure 10.7.: Distribution of NPV values from MC simulation

in a negative NPV. This is equivalent of 19.1% of the runs. For investment (in intervention measures) to be profitable (i.e. business justifiable) over the period of five years, the NPV must be positive. As such, probability of investment success is estimated at 81.9%.

10.7.1. Sensitivity Analysis

Sensitivity analysis studies the uncertainty in the output of a system and how such uncertainties are affected by different sources of uncertainties in the input variables. There are several methodologies to undertake sensitivity analysis, driven by type and characteristics of system under study and the input/output variables. In this study, changing *one-factor-at-a-time* method is used, to investigate what effect the change produces on output. In this method, changes are applied to one input variable while others are kept at their nominal values, and output variable is calculated. The input variable returns to its nominal value, then repeating the process for other input variables in the same way. It is assumed that the input variables are not correlated and no interaction exist between them (i.e. variables are independent) (Saltelli et al., 2009).

A sensitivity analysis is undertaken for the 10 most important input variables, listed in Table 10.14. The most important variables are selected based on their level of impact (positive or negative), and likelihood of occurrence.

Sensitivity analysis showed that none of the input variables are more than 1% sensitive, meaning that a change of 1% in the estimation, did not change the final average NPV or percentage of negative NPV values by 1%. Therefore,

Table 10.14.: Sensitivity analysis for the 10 most important factors

Input factor change	Percent Changed	Average NPV	Average NPV Change	% NPV Negative	% NPV Negative Change
No Change	0	9.11	0	19.16	0
Labour cost	1	9.05	0.06	19.2	0.04
Leakage maintenance	1	9.03	0.08	19.18	0.02
Emergency projects	1	9.13	0.02	19.25	0.09
Reputation	1	9.02	0.09	19.12	0.04
Environmental benefits	1	9.05	0.06	19.19	0.03
ODI reward/penalty	1	9.12	0.01	19.24	0.08
SIM score	1	9.16	0.05	19.14	0.02
Supply interruption	1	9.06	0.05	19.1	0.06
Customer service/ complaints	1	9.14	0.03	19.2	0.04
Optimising Capex use	1	9.16	0.05	19.12	0.038

the confidence in the estimations is relatively high, for if it is not completely accurate, the results would not fluctuate significantly.

10.8. Research business case and outcome exploitation

The concept of engineering doctorate is based on combining strength of both academia and the industry to deliver high quality practical research. The research is designed and undertaken to enhance the current knowledge in the area, while at the same time addressing the existing short and long term challenges within the industry. An important aspect of the research, therefore, is its commercial considerations, cost-saving and business benefits it can deliver, not only to the sponsoring entities but also to the sector and the wider society.

This section briefly highlights some of the benefits that calm network can provide and reviews potential business cases for exploitation of the research outcomes.

10.8.1. Research benefits

The project outcome is to enhance the system operation and improve assets' (particularly pipes') service lives. Any system performance that is impacted by the pipes' condition and structural integrity is therefore impacted by the research. These include asset management and renewal programme, leakage,

pipe repair teams, supply interruptions, water quality (discolouration) incidents, etc. There are also a number of indirect impacts of pipe failure, specifically relevant for failure in larger pipes, such as health issues, public and business disruptions, etc. Any insight into mitigating risk of such failures are of great importance and benefit to any water utility.

The project benefits are both short term and long term. Firstly, as a prevention exercise, enhanced insight to the network operation can assist addressing operational challenges in high risk areas. Moreover, high frequency pressure logging and the knowledge obtained from the data analytic, provide valuable tools for the system operatives to understand the system operational health. The real impact of the project outcome is its long term benefits. Inputs to the deterioration models, more efficient mains replacement and management programmes and mitigating risks of pipe failures are amongst such impacts.

Overseas benefit The project outcome can have potential values for communities outside the UK alike. Water networks with similar (physical and hydraulic) characteristics exist in many countries, indicating that such system face a similar phenomenon. Pipe failure is already a big problem in the US and various other countries, e.g. American Water Works Association (2012). The problem of dynamic loadings and impact of pipe failure is by no means limited to the UK, specifically as other countries are adopting intensified efficiency driven measures (e.g. more pump operations). Findings of the study have been discussed with representatives from water industry, both inside the UK and internationally (specifically Canada and Australia), and helped raising awareness in this area.

10.8.2. Exploiting the outcome

Alongside the project development, there were several cases where intermediate outcome of the project led to improvement in the network operation. These were mainly through exploiting the high frequency logging equipment and the analytic to develop understanding of operational challenges across the network and provide input for solution development.

Regarding technology readiness level, the data collection equipment and platform have evolved during the project (the hardware development has been part of another study undertaken partly in tandem with this research work). The analytical methods of data analysis, CPIS development and the statistical models are developed and ready to be implemented across the relevant case studies. A start-up company, Inflomatix, supported and partly-sponsored by

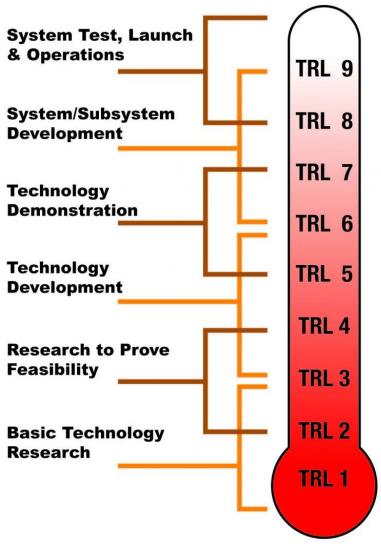


Figure 10.8.: Technology readiness stages-adopted from Mankins (2004)

Research and Innovation at Imperial College London is established to enhance the outcomes from the project findings (TRL 9 in Figure 10.8).

Resource needs Required resources to exploit the research outcome would depend on the scale of implementation. In general, the requirements are logging equipment, analytical tools (server, data communication technologies, analytical algorithm for hydraulic tracing, energy dissipation, cycle counting, CPIS development, statistical models, etc. and the man-power to undertake data collection and the analytical process. As the exercise involves several teams, high level of quality communication and liaison is needed. The process includes problem identification, liaising with operational teams to plan and execute the logging programme, review and analyse the collected data and other network information, solution development and implementation.

Risks and further research Further exploitation of the research will be through studying new parts of the network that requires additional local logging, analysis and intervention. The main risk to a successful exploitation of the project will be from lack of effective collaboration with the local teams, in both problem identification and solution development.

To enhance operational benefits, further research and development will be required to incorporate into the analytic and improve the statistical model with any new data collected. Studies into potential intervention measures could generate list of solutions that can be used to mitigate impacts of hydraulic instabilities. Currently, such measures are limited to soft-start on pumps

and possibly change of pump operation, and largely depending on the local knowledge. A library of potential solutions will not only be technically more suitable, but from practical point of view, it is easier and more cost effective to be implemented.

10.8.3. Knowledge transfer

The project was successful in disseminating its mission and outcomes across various levels of the sponsoring organisation and beyond. Due to its nature, interest and input from both operational and management teams were required for a successful delivery. Raising the awareness of both groups has helped shape several schemes and initiatives aiming to consider adverse impacts of pressure transient in the routine operation of the network as well as long term planning and management (e.g. deterioration models, mains planning and rehabilitation, etc.).

Externally, various other water utilities have been made aware of the work (e.g. through dissemination events, UKWIR workshop, etc.) and some have eagerly sought the outcome and taken steps to undertake similar paths. Within the water utilities, sharing the output could benefit operational and optimisation teams, asset creation and designers as well as leakage, pipe repair and maintenance teams. Potential ways to disseminate the project's outputs within a water utility are to:

- Directly work with local teams (operation, optimisation, asset creation, leakage, etc.) in order to highlight the direct benefit to their work
- Identify new implementation opportunities (e.g. with new schemes) to benefit from the project outcome and also disseminate the result to new teams
- Provide updates to relevant teams through successful case studies familiar to theirs
- Highlight the findings to the wider audience
- Presentations and general dissemination activities to wider areas of the business to raise awareness

10.9. Chapter summary

Adverse impacts of hydraulic instabilities on increasing pipe deterioration and ultimate failure across the water networks highlights the need for intervention

activities to prevent and mitigate such impacts. While detailed estimation for cost of pipe failure remains a challenge, a combination of cost data analysis, questionnaires, interviews, experts' opinions, companies' business plans and engineering judgement can provide reasonably reliable data to estimate total costs of pipe failure and costs associated with intervention activities.

In this chapter, a methodology is presented, using the concept of net present value combined with the MC simulation, in order to evaluate business justification of undertaking intervention measures to calm the network. The analyses illustrate that for DMAs impacted by hydraulic pressure variations, there is an 81.9% probability of profitability on investments made in prevention activities to mitigate these instabilities. Sensitivity analyses of the most important input factors show that the analyses and simulations are resistant to alterations to the factors' estimations.

The methodology provides a platform to evaluate cost effectiveness and business justifications for undertaking prevention and mitigation measures, specifically where associated costs and benefits are not precisely quantifiable. This methodology can be used for other network challenges and associated solutions and technologies, provided estimation of cost and benefit can be made.

Future work Lack of quality data and information about cost of pipe failure in water utilities continues to be the main challenge in evaluating profitability of any intervention measure. A programme of recording costs that arose from pipe burst events, specifically those with adverse impact beyond simple repair (labour and material), could provide valuable data for future business justification decisions. Any improvement to the quality and accuracy of the input cost data will enhance the confidence in the model output. The analysis can be extended by considering likelihood and significance of input variables (specifically qualitative ones), resulting in enhanced model reliability.

11. Conclusions and recommendations

The overall aim of the study was set out to enhance the understanding of the dynamic pressure variations in water supply networks and evaluate their impacts on pipe failures. Original objectives of the study, as listed in section 1.1.3 (and repeated here), were as following:

I. Provide unique insight into the dynamic hydraulic behaviour across the water supply networks and study the unsteady pressure variations and their characteristics, types, amplitudes, frequencies, energy wave dissipation mechanisms, etc.

II. Review the existing knowledge on the factors affecting and accelerating pipe failure mechanisms and critically examine the existing deterioration modelling approaches and their shortcomings

III. Design and execute an experimental programme to gather system representative high frequency pressure data across a large supply network

IV. Systematically investigate associations between unsteady and quasi-steady pressure variations and pipe's deterioration and failure at the network level, using the advances in statistical techniques

V. Develop a cost-benefit framework to evaluate cost effectiveness of achieving calm networks. The framework is to consider economical, sustainable and practical factors and the thresholds at which intervention actions aiming calm networks are justified

This study has fulfilled these objectives. It has argued that pipe structural degradation can be accelerated through adverse impact of stresses originated from the cyclic loadings of pressure variations. This research has explained the central importance of considering network's hydraulic characteristics (particularity quasi-steady and unsteady pressure changes) in network deterioration and benefits to be gained from calming networks.

11.1. Summary of findings

The most obvious finding to emerge from this study is that calming the network can offer significant savings by reducing risks of pipe failures and consequential network challenges.

The findings enhance our understanding of occurrence and characteristics of the rapid pressure changes and hydraulic instabilities across water networks. One of the more significant findings to emerge from this study is that the extend of pressure fluctuations across the network is often much more severe than realised by the practitioners and network managers. There are growing efforts to keep the networks in the steady state condition, however, efficiency driven and energy-saving measures often result in higher levels of interventions and operational alterations (for instance more aggressive pump and valve operations). These, in turn, lead to growing levels of unsteady pressure fluctuations. Significant pressure logging undertaken in this study illustrated the extend that water supply networks are exposed to pressure fluctuations. Most system experience one or several types of pressure fluctuations events, either frequent small-amplitude pressure events, large-amplitude events with low frequency or a combination of these.

In pipeline design and management processes, often, only individual large-amplitude (positive or negative) pressure transient events are considered. While these are important, the findings from this study suggest that cumulative damage from the frequent cyclic loadings can accelerate pipe deterioration. Multiple logistic regression analysis revealed associations between probability of pipe failures and level of hydraulic instability in the pipelines.

Pipe failure is a complex phenomenon impacted by numerous factors. The current findings add to a growing body of literature on pipe failure investigations. The study confirms previous findings and contributes additional evidence on associations between pipe characteristics, such as length, material and diameter, and the rate of failure. In addition to the factors conventionally known to be affecting pipe failures, the relevance of CPIS to increased pipe failure probabilities is clearly supported by the current findings.

Whilst this study did not implement inclusion of CPIS in deterioration models, it did substantiate importance of considering this metric as a significant factor in the future model development. The evidence from this study suggests that statistical deterioration modelling approaches can be enhanced by taking into consideration role of cumulative damages from system's hydraulic.

Taking into account the study's sampling programme, this research can serve as a base for future studies in design and execution of large-scale data

collection in water supply studies.

11.2. Novelty and key contributions

The findings from this study make several contributions to the current literature.

First, the research developed a framework to systematically investigate the impact of the hydrodynamic conditions and evaluate benefits of calm networks. A large-scale sampling programme was designed and executed to collect system representative high frequency pressure data. This provided unique insights into the network's hydraulic operations, highlighting the significance of pressure variations. A major novelty of the study was the investigation and validation of the cumulative pressure induced stress through cycle counting of the pressure variations. Associations between pipe failures and factors contributing to pipe failure acceleration were systematically investigated. New enhanced statistical models were developed to study this, particularly to evaluate and quantify impacts of CPIS. Another major contribution from this study is a methodology developed to evaluate cost-effectiveness of innovative intervention activities aiming to achieve calm networks.

Additional contributions from this study included developing simplified hydraulic tracing and energy dissipation methods to estimate the energy level across the network from limited scattered data points. A pipe failure data collection platform was developed in this study than can contribute to improved data quality for future investigations. Furthermore, recommendations are provided for level of instrumentation for studying the system's dynamic hydraulic and inclusion of CPIS in pipe deterioration models.

The industry will benefit from this work as it provides a deeper understanding of the causes of pipe failure and in doing so provides a tool for management and avoidance of bursts. This will impact water distribution network management and control, investment planning, pipe replacement planning and prioritisation and capital investment planning. The benefits of improved understanding of the network, the ability to calm and control the network, to improve expenditure profiling and most importantly to avoid even a small number of (catastrophic) bursts are Significant.

11.3. Recommendations for future works

Below are several recommendations for future research to be undertaken in this area:

Pipes condition assessment Existing micro-cracks, flaws and other imperfections on pipes can significantly accelerate fatigue failure process. Coupled with the CPIS metric, high quality pipe condition assessment data can enhance the pipe health investigation and predicted probability of failure. Existing pipe condition assessment methods tend to focus on high risk large diameter pipes, often at considerable cost, therefore mostly limited to a short section of the pipeline. The obvious route for future research would be to develop new condition assessment methods, able to cover larger parts of the network more efficiently. For smaller mains, and for a wide-spread network coverage, pipe assessment data *availability* is arguably a more urgent need than the data *accuracy*. As such, a reasonable approach to tackle the lack of data could be new data-driven methods that can be developed to estimate pipe's conditions from limited observations, at a reasonable level of confidence.

CPIS improvement This study is a start in developing our understanding of the pressure-induced cumulative stress and damage. The concept can benefit from future research in a number of ways. The algorithm to estimate CPIS can be enhanced by providing special treatments for unusual events (for instance unusually large events that can have additional deteriorating impact on the pipes). Moreover, the CPIS development can be improved by considering pipe conditions and effect of micro-cracks, notches, etc. in accelerating fatigue failures. Increasing high frequency data collection deployment can increase confidence in the CPIS estimations across the network and in turn enhance the statistical models that use these metrics.

Deterioration models An implication of the research findings is that CPIS, representing pipes' hydraulic health, should be taken into account in asset management strategies, pipe replacement programmes and operational improvement plans. Further research regarding implementation of the CPIS in deterioration models can provide great acceleration for gaining benefits from the calm networks.

Hydraulic tracing and energy dissipation Additional developments in hydraulic tracing can provide significant benefit to a wide variety of future studies, where hydraulic relationships (e.g. shortest distance) between network nodes are of interest. Most existing methods, mainly developed for investigating water quality topics, often lack the accuracy expected in other research studies, for instance when calculation of the shortest hydraulic path is of interest.

Currently an active area of research, improvement in our understanding of the

energy dissipation mechanisms, wave propagation and energy transformation across the network can enhance the energy level estimation across the network. This is particularly beneficial when large-scale studies across the network are undertaken.

Future data collections A significant emphasise should be given to improving quality and quantity of data across the industry. Various decisions on the system's operation and management are relying on the outcomes of tools and analyses that are developed based on such data. Particularly, improved data collection and management should be considered for pipe failure records, operational activities, network operational changes and network assets. Improvement should be made on both existing datasets and also in planning for new data collections. Often, with sufficient planning, vital data can be collected as by-products of other activities, occurring minimal costs. For instance, collecting data on the buried pipes when a repair activity is taking place is a cost effective method of enhancing the quality of underground asset datasets. Table 11.1 provides a list of suggested data and information that should be collected, in order to develop more reliable datasets for future investigations on pipe failures and system hydraulics.

Pipe failure cost estimate An important area for future work, potentially more targeted at the industry than academia, is to develop datasets that can assist in estimating total cost of pipe failure. While certain elements of such estimates might be case-specific, in general, having some estimates of such costs can significantly assist driving the research needs and innovations across the industry. Recommendations on the required data are provided in chapter 10.

Library of case studies As the area of unsteady pressure changes and pressure transient enjoy new popularity in research and industrial contexts, provision of a library of datasets, representing various types of hydraulic instabilities (e.g. event characteristics, amplitude, frequency, rate of change, event distribution and change over time, etc.) can act as a base for future studies. New methods can be tested on a common test cases, providing improvement to the comparative investigations.

Table 11.1.: Information and data to be collected for enhancing future analyses

Date for	Data attribute
Assets	Main's material and size Main's external condition (corrosion level) Main's external/cathodic protection Joint type Type, made and model of upstream pump Type, made and model of upstream valve
Operation and hydraulic	Supply arrangement and operational regime of the DMA and supply zone High frequency pressure logging Operation regime and performance of upstream pump Operation regime and performance of upstream and nearby valves Nearby large industrial users
Burst and repaired pipe	Burst and repair dates Burst location information (coordinates, land use, etc.) Burst type Burst detection method Repair type Replacement main's material, length and size Pipe's sample information
Surrounding environment	Condition of ground surrounding the pipe (soil corrosivity and fracture potential) Bedding under the pipe Nearby services Surface use
Photographic survey	Repair site Leaking and repaired main

References

Ahammed, M. and Melchers, R. (1995), 'Probabilistic analysis of pipelines subjected to pitting corrosion leaks', *Engineering Structures* **17**(2), 74–80.

Ambrose, M., Burn, S., DeSilva, D. and Rahilly, M. (2008), Life cycle analysis of water networks, *in* 'Plastics Pipe XIV: Plastics Pipes Conferences Association', Vol. 22.

American Water Works Association (2001), Dawn of the replacement era-Reinvesting in drinking water infrastructure, Technical report, Denver.

American Water Works Association (2003), *Ductile-iron pipe and fittings*.

American Water Works Association (2012), *Buried no longer-confronting Americas water infrastructure challenge*, Water Industry Technical Action Fund.

Andreou, S., Marks, D. and Clark, R. (1987), 'A new methodology for modelling break failure patterns in deteriorating water distribution systems-Applications', *Advances in Water Resources* **10**(1), 11–20.

Arshad, A. (2012), 'Net Present Value is better than Internal Rate of Return', *Interdisciplinary Journal of Contemporary Research in Business* **4**, 211–219.

Ascher, H. and Feingold, H. (1984), *Repairable systems reliability*.

ASTM (2011), 'Standard practices for cycle counting in fatigue analysis - (ASTM_e_1049_85_2011-e1) - American Society for Testing and Materials'.

Atkinson, K., Whiter, J., Smith, P. and Mulheron, M. (2002), 'Failure of small diameter cast iron pipes', *Urban Water* **4**(3), 263–271.

Baboian, R. (2005), *Corrosion tests and standards: application and interpretation*, American Society for Testing and Materials, Philadelphia, PA.

Bannantine, J., Comer, J. and Handrock, J. (1990), *Fundamentals of metal fatigue analysis*, Prentice Hall.

Belmonte, H., Mulheron, M. and Smith, P. (2009), 'Some observations on the strength and fatigue properties of samples extracted from cast iron water mains', *Fatigue and Fracture of Engineering Materials and Structures* **32**(11), 1–10.

Bhat, S. and Patibandl, R. (2011), Metal fatigue and basic theoretical models: a review, *in* E. Morales, ed., 'Alloy Steel - Properties and Use', InTech.

Box, G., Hunter, S. and Hunter, W. (2005), *Statistics for experimenters: design, innovation, and discovery*, Wiley-Interscience.

Boxall, J., O'Hagan, A., Pooladsaz, S. and Saul, A. (2007), 'Estimation of burst rates in water distribution mains', *Proceedings of the ICE - Water Management* **160**(2), 73–82.

Brunone, B., Karney, B., Mecarelli, M. and Ferrante, M. (2000), 'Velocity profiles and unsteady pipe friction in transient flow', *Journal of Water Resources Planning and Management* **126**(4), 236–244.

Cabinet Office (2013), 'Discount rates and net present value'.
URL:
https://data.gov.uk/sib˙knowledge˙box/discount-rates-and-net-present-value

Camarinopoulos, L., Chatzoulis, A., Frontistou-Yannas, S. and Kallidromitis, V. (1999), 'Assessment of the time-dependent structural reliability of buried water mains', *Reliability Engineering and System Safety* **65**(1), 41–53.

Chaudhry, M. H. (1987), *Applied hydraulic transients*, Van Nostrand Reinhold.

Cima, J. and Peters, D. (2009), National sewers and water mains failure database, Technical report, UK Water Industry Research Limited.

CIPRA (1927), *Handbook of cast iron pipe: for water, gas, sewerage, and industrial service*, Cast Iron Pipe Research Association, Chicago.

Cohen, J., Cohen, P., West, S. and Aiken, L. (2013), *Applied multiple regression/ correlation analysis for the behavioral sciences*, Routledge.

Collins, R. (2015), 'High resolution pressure logging in water supply networks-interview'.

Cooper, N., Blakey, G., Sherwin, C., Ta, T., Whiter, J. and Woodward, C. (2000), 'The use of GIS to develop a probability-based trunk mains burst risk model', *Urban Water* pp. 1–7.

Covas, D., Stoianov, I. I., Mano, J. F., Ramos, H., Graham, N. and Maksimovic, C. (2005), 'The dynamic effect of pipe-wall viscoelasticity in hydraulic transients. Part II-model development, calibration and verification', *Journal of Hydraulic Research* **43**(1), 56–70.

Cranfield University (1997), 'The national soil map and soil classification by Soil Survey and Land Research Centre'.

Creasey, J. and Lynn, S. (2011), Impact of valve operations on network performance, Technical Report P8560.02, Water Research Centre, Swindon, UK.

Cromwell, J. (2002), *Costs of infrastructure failure*, American Water Works Association.

Davis, J. (2004), *Tensile testing*, ASM International.

de Oliveira, D. P. (2010), Spatial data analysis of networked infrastructure failure data An application for condition assessment of WDS, PhD thesis.

Denk, M. and Bakran, M. (2013), Comparison of counting algorithms and empiric lifetime models to analyze the load-profile of an IGBT power module in a hybrid car, *in* '3rd International Electric Drives Production Conference, 2013', IEEE, pp. 1–6.

DNVGL (2014), 'Synergi Water - Advancing smart water networks'.
URL: *https://www.dnvgl.com/services/synergi-water-modules-4123*

Doleac, M., Lackey, S. and Bratton, G. (1980), Prediction of time-to failure for buried cast iron pipe, *in* 'American Water Works Association Annual Conference'.

Domencich, T. and McFadden, D. (1975), *Urban travel demand - a behavioral analysis*, North-Holland Publishing Co.
URL: *https://trid.trb.org/view.aspx?id=48594*

Dowling, N. (1998), *Mechanical behaviour of materials: engineering methods for deformation, fracture, and fatigue*, 2nd edn, Prentice Hall.

Downing, S. and Socie, D. (1982), 'Simple rainflow counting algorithms', *International Journal of Fatigue* **4**(1), 31–40.

Duan, H.-F., Ghidaoui, M., Lee, P. and Tung, Y.-K. (2010), 'Unsteady friction and visco-elasticity in pipe fluid transients', *Journal of Hydraulic Research* **48**(3), 354–362.

Enders, C. K. (2010), *Applied missing data analysis*, Guilford Press.

Engrasp, A. (2008), 'Engineering toolbox stress-life fatigue analysis module'.
URL: *http://www.fea-optimization.com/etbx/stresslife'help.html*

Ford, C. (2016), 'Visualizing the effects of logistic regression'.
URL: *http://data.library.virginia.edu/visualizing-the-effects-of-logistic-regression/*

Fox, J. (1987), *Effect displays for generalized linear models*, York University Toronto - Institute for Social Research.
URL: *http://www.jstor.org/stable/pdf/271037.pdf*

Fox, J. (2003), 'Effect displays in R for generalised linear models', *Journal of Statistical Software* **8**(15), 1–27.
URL: *http://psfaculty.ucdavis.edu/bsjjones/effectdisplays.pdf*

Fox, J. and Andersen, R. (2006), 'Effect displays for multinomial and proportional-odds logit models', *Sociological Methodology* **36**(1), 225–255.

Fuller, W. (2011), *Sampling statistics*, John Wiley & Sons.

Gaewski, P. and Blaha, F. (2007), Analysis of total cost of large diameter pipe failures, *in* 'Proceeding of AWWA Research Symposium Distribution Systems: the Next Frontier', American Water Works Association.

Ginzel, R. and Kanters, W. (2002), Pipeline corrosion and cracking and the associated calibration considerations for same side sizing applications, Technical report, Eclipse Scientific Products Inc.

Goulter, I., Davidson, J. and Jacobs, P. (1993), 'Predicting water main breakage rates', *Journal of Water Resources Planning and Management* **119**(4), 419–436.

Grigg, N. (2007), *Main break prediction, prevention, and control*, American Water Works Association.

Halvarsson, D. (2013), 'Firm dynamics-the size and growth distribution of firms'.
URL: *http://www.diva-portal.org/smash/record.jsf*

Hansen, M., Hurwitz, W. and Madow, W. (1993), *Sample survey methods and theory*, Vol. 1, Wiley.

Hembara, O. and Andreikiv, O. (2012), 'Effect of hydrogenation of the walls of oil-and-gas pipelines on their soil corrosion and service life', *Materials Science* **47**(5), 598–607.

Herrera, M., Abraham, E. and Stoianov, I. (2015), Graph-theoretic surrogate measures for analysing the resilience of water distribution networks, *in* 'Procedia Engineering', Vol. 119 of *Computing and Control for the Water Industry*, pp. 1241–1248.

Hertzberg, R. (1996), *Deformation and fracture mechanics of engineering materials*, John Wiley & Sons.

Hoskins, A. (2015), Monitoring the hydraulic dynamics in water distribution systems, PhD thesis, Imperial College London.

Hosmer, D. W., Lemeshow, S. and Sturdivant, R. X. (2013), *Applied logistic regression*, 3rd edn, John Wiley & Sons.

Jacobs, P. and Karney, B. (1994), GIS development with application to cast iron water main breakage rate, *in* '2nd International Conference of Water Pipeline Systems'.

Jarvis, N., Hollis, J., Nicholls, P., Mayer, T. and Evans, S. (1997), 'MACRO-DB a decision-support tool for assessing pesticide fate and mobility in soils', *Environmental Modelling and Software* pp. 251–265.

Jianhua, L. and Sveinung, S. (1998), 'Statistical approach for describing failures and lifetimes of water mains', *Water Science and Technology* **38**(6), 209–217.

Jonkman, A. and Wong, A. (2001), 'Method helping to quantify benefits of applying new upstream technology', *Oil and Gas Journal* **99**(11), 34 – 36.

Karney, B. (1990), 'Energy relations in transient closed-conduit flow', *Journal of Hydraulic Engineering* **116**(10), 1180–1196.

Karney, B. and Filion, Y. (2003), Energy dissipation mechanisms in water distribution systems, *in* '4th Joint Fluids Summer Engineering Conference', pp. 2771–2778.

Karney, B., Malekpour, A. and Nault, J. (2014), 'An energy approach to studying pipe network transients', *Procedia Engineering* **89**, 1298–1305.

Kettler, A. and Goulter, I. (1985), 'An analysis of pipe breakage in urban water distribution networks', *Canadian Journal of Civil Engineering* **12**(2), 286–293.

Kiefner, J., Maxey, W., Eiber, R. and Duffy, A. (1973), Failure stress levels of flaws in pressurized cylinders, *in* 'Progress in Flaw Growth and Fracture Toughness Testing', ASTM International.

Kirmeyer, G., Richards, W. and Smith, C. (1994), *Assessment of water distribution systems and associated research needs*, American Water Works Association.

Kleiner, Y. and Rajani, B. (2001), 'Comprehensive review of structural deterioration of water mains: statistical models', *Urban Water* **3**(3), 131–150.

Kleiner, Y., Rajani, B. and Sadiq, R. (2006), 'Failure risk management of buried infrastructure using fuzzy-based techniques', *Journal of Water Supply: Research and Technology-AQUA* **55**(2), 81–94.

Lambert, A. (2001), What do we know about pressure-leakage relationships in distribution systems, *in* 'IWA Conference in Systems Approach to Leakage Control and Water Distribution System Management', Citeseer, Brno, Czech Republic.

Lambert, A. and Fantozzi, M. (2010), Recent developments in pressure management, *in* 'International Water Association Specialised Conference', Sao Paolo, Brazil.

Lampka, J. (2016), 'Monte Carlo Simulation'.
URL: *http://www.financeisland.com/tutorials/MonteCarlo*

Lansey, K. and Mays, L. W. (2000), Hydraulics of water distribution systems, *in* 'Water Distribution Systems Handbook', McGraw-Hill education, p. 912.

Lee, Y.-L., Hathaway, R. and Pan, J. (2005), *Fatigue testing and analysis: theory and practice*, Butterworth-Heinemann, Amsterdam.

Leonard, B. (2013), 'The Science of nondestructive testing'.
URL: *https://www.cnde.iastate.edu/nde-education/*

Lin, W.-H. and Hauptmann, A. (2004), Merging rank lists from multiple sources in video classification, *in* 'IEEE Multimedia International Conference', Vol. 3, IEEE, pp. 1535–1538.

Long, S. and Freese, J. (2006), *Regression models for categorical dependent variables using Stata*, 2nd edn, Stata Press.

Louviere, J., Hensher, D. and Swait, J. (2000), *Stated choice methods: analysis and applications*, Cambridge University Press.

MacKellar, S. (2003), Nationally agreed failure data and analysis methodology for water mains-volume 2: protocol for the capture and reporting of data on mains failures, Technical report, UK Water Industry Research Limited.

Makar, J., Desnoyers, R. and McDonald, S. (2001), Failure modes and mechanisms in gray cast iron pipe, *in* 'International Conference on Underground Infrastructure Research', Citeseer, pp. 303–312.

Mankins, J. (2004), 'Technology readiness levels - a white paper, Office of space access and technology-NASA'.

Markl, A. and Louisville, K. (1955), Piping flexibility analysis, The American Society of Mechanical Engineers, pp. 127–149.

Marshall, P. (2001), The residual structural properties of cast iron pipes-structural and design criteria for linings for water mains, Technical Report 01/WM/02/14, UK Water Industry Research Limited.

MathWorks (2015), 'MATLAB, 2014a, The MathWorks Inc., Massachusetts'.

Maurer, M. (2013), 'Network deterioration modelling - Lecture series at ETH Zurich'.

McFadden, D. (1973), Conditional logit analysis of qualitative choice behavior, *in* 'Frontiers in Econometrics', Wiley, New York.
URL: *https://elsa.berkeley.edu/reprints/mcfadden/zarembka.pdf*

McFadden, D. (1977), *Quantitative methods for analyzing travel behavior of individuals: some recent developments*, Institute of Transportation Studies, University of California.
URL: *http://core.ac.uk/download/pdf/6448852.pdf*

McLaney, E. and Atrill, P. (2015), *Accounting and finance-an introduction*, 8th edn, Pearson Higher Education.

McMullen, L. (1982), Advanced concepts in soil evaluation for exterior pipeline corrosion, *in* 'Proceedings of the AWWA Annual Conference', American Water Works Association.

Meniconi, S., Brunone, B., Ferrante, M. and Massari, C. (2014), 'Energy dissipation and pressure decay during transients in viscoelastic pipes with an in-line valve', *Journal of Fluids and Structures* **45**, 235–249.

Mohebbi, H., Jesson, D., Mulheron, M. and Smith, P. (2010), 'The fracture and fatigue properties of cast irons used for trunk mains in the water industry', *Materials Science and Engineering* **527**(21-22), 5915–5923.

Murakami, Y., ed. (1992), *The rainflow method in fatigue: the Tatsuo Endo memorial volume*, Butterworth-Heinemann.

Newell, B. (2016), 'The drive to innovation - a methodology for quantifying costs and benefits of innovative technologies in the UK water supply and distribution sector - MSc thesis - Imperial College London'.

Nishiyama, M. and Filion, Y. (2013), 'Review of statistical water main break prediction models', *Canadian Journal of Civil Engineering* **40**(10), 972–979.

Ofwat (2015), Service incentive mechanism-guidance for collating customer service information for calculating the SIM score.
URL: *http://www.ofwat.gov.uk/wp-content/uploads/2015/11/gud˙pro201503sim.pdf*

Ofwat (2016), 'Outcome delivery incentive rates for Severn Trent Water'.
URL: *https://www.ofwat.gov.uk/wp-content/uploads/2015/10/pap˙tec201503odimenusvt.pdf*

Paradkar Brahmanand, A. (2012), An evaluation of failure modes for cast iron and ductile iron water pipes, PhD thesis, University of Texas at Arlington.
URL: *https://uta-ir.tdl.org/uta-ir/handle/10106/11660*

Pritchard, O., Hallett, S. and Farewell, T. (2013), Soil corrosivity in the UK impacts on critical infrastructure, Technical report, Cranfield.

Raine, D. (2014), 'Evolution of the network hydraulic characteristics in Severn Trent Water East region'.

Rajani, B. and Kleiner, Y. (2001), 'Comprehensive review of structural deterioration of water mains: physically based models', *Urban Water* **3**(3), 151–164.

Rajani, B. and Kleiner, Y. (2004), 'Nondestructive inspection techniques to determine structural distress indicators in water mains', *Evaluation and Control of Water Loss in Urban Water Networks* pp. 1–20.

Rajani, B. and Kleiner, Y. (2011), Fatigue failure of large-diameter cast iron mains, *in* 'Water Distribution System Analysis', Tucson, AZ, USA, pp. 1146–1159.

Rajani, B. and Makar, J. (2000), 'A methodology to estimate remaining service life of grey cast iron water mains', *Canadian Journal of Civil Engineering* **27**(6), 1259–1272.
URL: *http://www.nrcresearchpress.com/doi/abs/10.1139/l00-073*

Rajani, B. and Tesfamariam, S. (2004), 'Uncoupled axial, flexural, and circumferential pipe soil interaction analyses of partially supported jointed water mains', *Canadian Geotechnical Journal* **41**(6), 997–1010.

Rajani, B., Zhan, C. and Kuraoka, S. (1996), 'Pipe soil interaction analysis of jointed water mains', *Canadian Geotechnical Journal* **33**(3), 393–404.

Rajeev, P., Kodikara, J., Robert, D., Zeman, P. and Rajani, B. (2013), Factors contributing to large diameter water pipe failure as evident from failure inspection, *in* 'Strategic Asset Management of Water and Wastewater Infrastructure: Leading Edge Strategic Asset Management', Sydney, Australia.

Randall-Smith, M., Russell, A. and Oliphant, R. (1992), *Guidance Manual for the Structural Condition Assessment of Trunk Mains*, Water Research Centre.

Rathnayaka, S., Shannon, B., Vahid, S., Guo, M., Sinclair, D. and Kodikara, J. (2016), Reducing severity of pressure transients in water supply networks, *in* 'OzWater: Australia's International Water Conference'.

RCoreTeam (2014), 'R: A language and environment for statistical computing'.
URL: *http://www.R-project.org/*

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

Rice, M. and Harris, G. (2005), 'Comparing effect sizes in follow-up studies: ROC Area', *Law and Human Behavior* **29**(5), 615–620.

Roberge, P. (1999), *Handbook of corrosion engineering*, McGraw-Hill Professional.

Romer, A., Bell, G., Duranceau, S. and Foreman, S. (2005), *External corrosion and corrosion control of buried water mains*, American Water Works Association.

Rosenfeld, M. and Kiefner, J. (2006), Basics of metal fatigue in natural gas pipeline systems - a primer for gas pipeline operators, Technical report, Technology for Energy Pipelines, Houston, Texas.

Rychlik, I. (1993), 'Note on cycle counts in irregular loads', *Fatigue and Fracture of Engineering Materials and Structures* **16**(4), 377–390.

Saltelli, A., Chan, K. and Scott, E. (2009), *Sensitivity analysis*, Wiley.

Savic, D. and Banyard, J., eds (2011), *Water distributions systems*, ICE Publishing, London.

Severn Trent Water (2015), 'Better value, better services, and a healthier environment: Severn Trent Water business plan 2015-2020'.

Severn Trent Water Asset Strategy (2015), 'Interviews on estimating total cost of pipe failure'.

Shamir, U. and Howard, C. (1979), 'An analytic approach to scheduling pipe replacement', *Journal of American Water Works Association* **71**(5), 248–258.

Silva-Araya, W. F. and Chaudhry, M. H. (1997), 'Computation of energy dissipation in transient flow', *Journal of Hydraulic Engineering* **123**(2), 108–115.
URL: *http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9429(1997)123:2(108)*

Soil Pipe Institute, C. I. (2006), *Cast iron soil pipe and fittings handbook*, 12th edn, Tennessee.

Stoianov, I. and Sadowski, A. (2015), Forensic investigation of a failure of a cast iron pipe section, Technical report, Imperial College London.

Stuart, J. (2011), Bursts per 1000km project report, Technical report, Severn Trent Water.

Tape, T. (2016), 'Interpreting diagnostic tests - ROC curves'.
URL: *http://gim.unmc.edu/dxtests/roc3.htm*

Thompson, S. K. (2012), *Sampling*, John Wiley & Sons.

Thorley, A. (2004), *Fluid transients in pipeline systems: a guide to the control and suppression of fluid transients in liquids in closed conduits*, 2nd edn, Professional Engineering Publishing.

Thornton, J. and Lambert, A. (2005), Progress in practical prediction of pressure: leakage, pressure: burst frequency and pressure: consumption relationships, *in* 'Proceedings of International Water Association Special Conference on Leakage', pp. 12–14.

United Nations (2015), 'World population prospects: the 2015 revision, key findings and advance tables.'.

Vipulanandan, C., Qiao, W. and Hovsepian, H. (2011), Case studies on water pipeline failures in the active zone, *in* 'Geo-Frontiers', American Society of Civil Engineers, pp. 2474–2483.

Walski, T., Daviau, J. and Coran, S. (2004), Effect of skeletonization on transient analysis results, *in* 'Critical Transitions in Water and Environmental Resources Management', American Society of Civil Engineers, pp. 1–10.

Water Research Foundation (2012), What pipe breaks and leaks reveal about pipe health, Technical report.

Wilkie, J. (2014), 'Changes to the hydraulics and operation of water distribution networks and operating assets'.

Wilks, S. S. (1932), 'Moments and distributions of estimates of population parameters from fragmentary samples', *The Annals of Mathematical Statistics* **3**(3), 163–195.
URL: *http://www.jstor.org/stable/2957624*

Wylie, B. (1978), *Fluid transients*, McGraw-Hill Inc.

Wylie, B., Streeter, V. and Suo, L. (1993), *Fluid transients in systems*, Prentice Hall.

Zadeh, L. A. (1994), 'Fuzzy logic, neural networks, and soft computing', *Communications of the Association for Computing Machinery* **37**(3), 77–85.

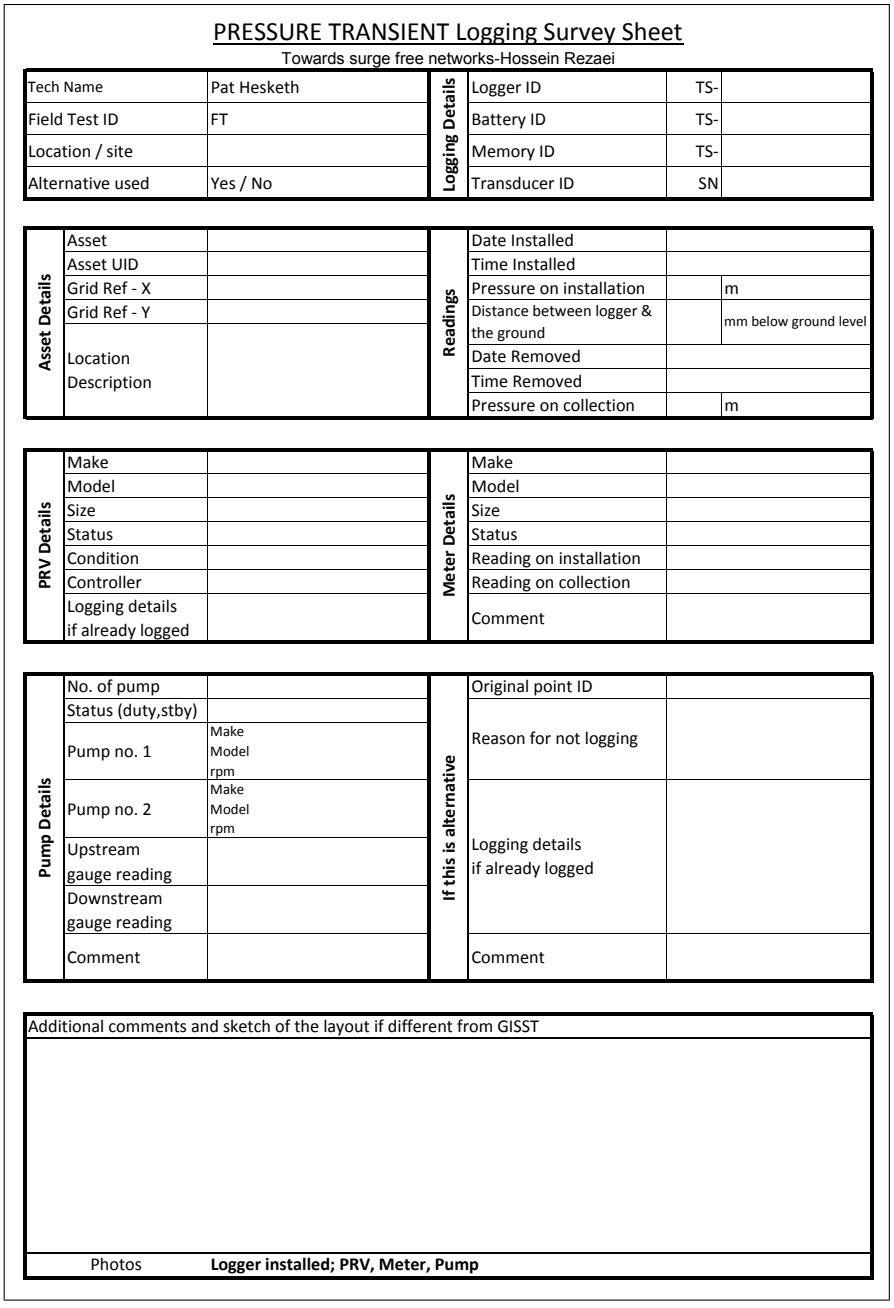
Zarembka, P. (1974), *Frontiers in econometrics*, Academic Press.

Appendices

A. Experimental programme

A.1. PT data collection form

Data collection form for high frequency pressure data collection.



PRESSURE TRANSIENT Logging Survey Sheet

Towards surge free networks-Hossein Rezaei

Tech Name	Pat Hesketh	Logging Details	Logger ID	TS-	
Field Test ID	FT		Battery ID	TS-	
Location / site			Memory ID	TS-	
Alternative used	Yes / No		Transducer ID	SN	

Asset Details	Asset		Readings	Date Installed		
	Asset UID			Time Installed		
	Grid Ref - X			Pressure on installation		m
	Grid Ref - Y			Distance between logger & the ground		mm below ground level
	Location Description			Date Removed		
				Time Removed		
				Pressure on collection		m

PRV Details	Make		Meter Details	Make	
	Model			Model	
	Size			Size	
	Status			Status	
	Condition			Reading on installation	
	Controller			Reading on collection	
	Logging details if already logged			Comment	

Pump Details	No. of pump		If this is alternative	Original point ID	
	Status (duty,stby)			Reason for not logging	
	Pump no. 1	Make Model rpm			
	Pump no. 2	Make Model rpm		Logging details if already logged	
	Upstream gauge reading				
	Downstream gauge reading				
	Comment			Comment	

Additional comments and sketch of the layout if different from GISST
Photos **Logger installed; PRV, Meter, Pump**

Figure A.1.: Pressure transient data collection form

A.2. Burst data collection

Examples from burst data collection exercise, described in section A.2, chapter 3. Collected information confirms a number of issues discussed in chapter 2. For instance, loss of wall thickness as the result of corrosion is illustrated in Figure A.2. Furthermore, pipes' surrounding soils often contains external objects (e.g. sizeable rocks, tree roots, etc. shown in Figures A.3 and A.4) imposing additional stress on the pipe and creating stress concentration points.



Figure A.2.: Examples of corrosion and wall thickness loss



Figure A.3.: Stress concentration from external objects (tree roots)



Figure A.4.: Stress concentration from external objects (rocks)

A.2.1. Tablet application

Screen-capture of tablet application used for burst data collection.

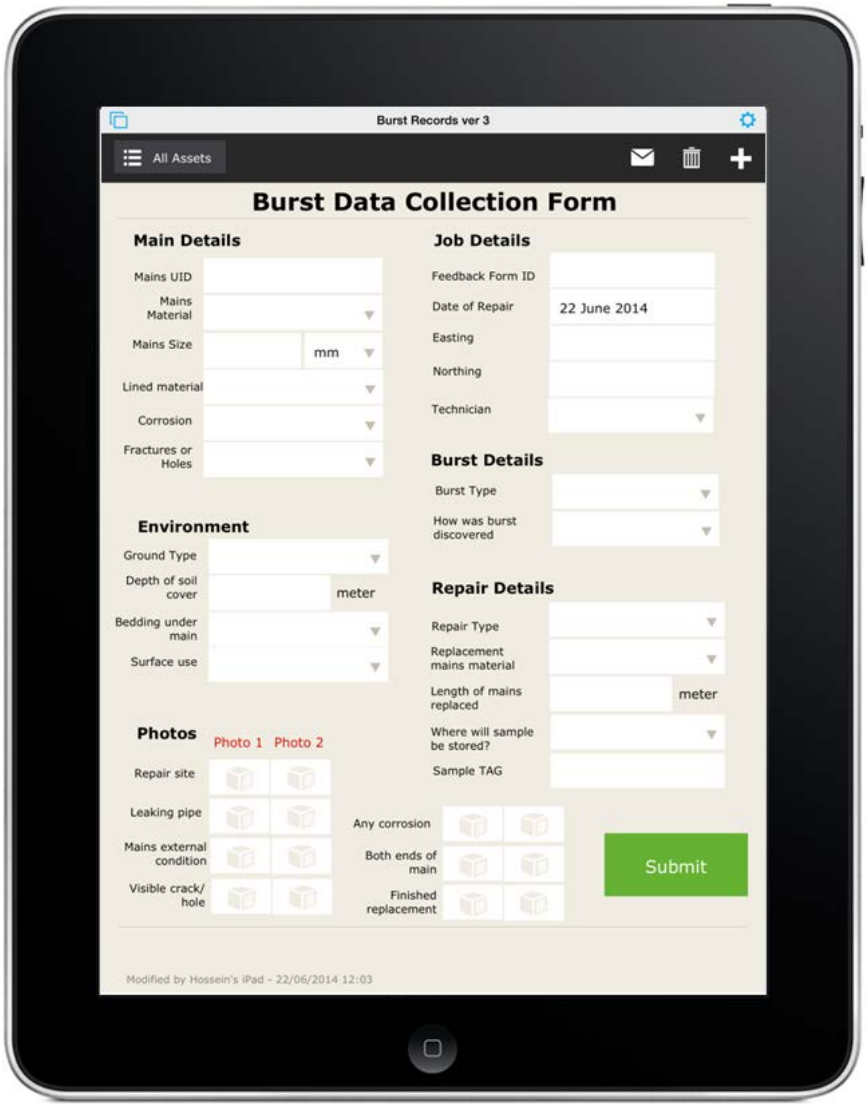


Figure A.5.: Burst data collection tablet application

B. Statistical analysis

B.1. Soil Fracture Potential (shrink/swell) level

Table B.1.: Soil shrink/swell level, adopted from Jarvis et al. (1997)

Shrink/ Swell ID	Shrink/ Swell Ref.	Shrink/ Swell Level	Soil Condition
A	1	very low	Hard rock, gravel, or sandy or coarse loamy soil
B	2	low	Heavy loam textures
C	3	moderate	Clayey overlying non-swelling slay, shale or marl
D	4	high	Clayey overlying swelling or lake clay
E	5	very high	Clayey overlying brownish swelling clay
D(D-E)	6	high*	Alluvial clay or peat with very high shrink-swell potential that is realized with drainage to >2m

B.2. Soil corrosivity level

Table B.2.: Soil corrosivity level, adopted from Jarvis et al. (1997)

Corrosivity ID	Corrosivity Ref.	Corrosivity Level	Soil Condition
A	1	Non-aggressive	Deep, permeable, well drained
B	2	Slightly Aggressive	Occasional wet, permeable subsoil or slightly seasonal water logging in slowly permeable soils
C	3	Moderately Aggressive	Acid; permeable soils with occasionally wet subsoil or permeable soils seasonally affected by groundwater; or dense, clay-rich, slowly permeable, seasonally waterlogged soils
D	4	Highly Aggressive	Moderately permeable, clay-rich or acid soils with seasonally fluctuating groundwater; or deep peat
E	5	Very highly Aggressive	Deep, acid, raw peat; or marine alluvial soils affected by fluctuating groundwater
Rock	6	Impermeable Rock	Shallow soils over massive bedrock where excavations may create a sump
C(B-D)	3*	Slightly to Highly Aggressive (sulphate)	Acid; permeable soils with occasionally wet subsoil or permeable soils seasonally affected by groundwater; or dense, clay-rich, slowly permeable, seasonally waterlogged soils. These soils will be moderately aggressive (3) if only a trace of sulphate is present. However if no sulphates are present, these will only be slightly aggressive (2) while if there is more than a trace of sulphates, these soils will be highly aggressive (4).
D(C-E)	4*	Moderately to Very Highly Aggressive (sulphate)	Moderately permeable, clay-rich or acid soils with seasonally fluctuating groundwater; or deep peat. These soils will be highly aggressive (4) if only a trace of sulphate is present. However if no sulphates are present, these will only be moderately aggressive (3) while if there is more than a trace of sulphates, these soils will be very highly aggressive (5).
E(D-E)	5*	Highly to Very Highly Aggressive (sulphate)	Deep, acid, raw peat; or marine alluvial soils affected by fluctuating groundwater. These soils will be very highly aggressive (5) if any sulphate is present. However if no sulphates are present, these will only be highly aggressive (4).

B.3. Pipe segment length

Figure B.1 shows the distribution of pipe segments' length. The majority of pipe segments are short (<10 m), with a similar distribution across different pipe materials.

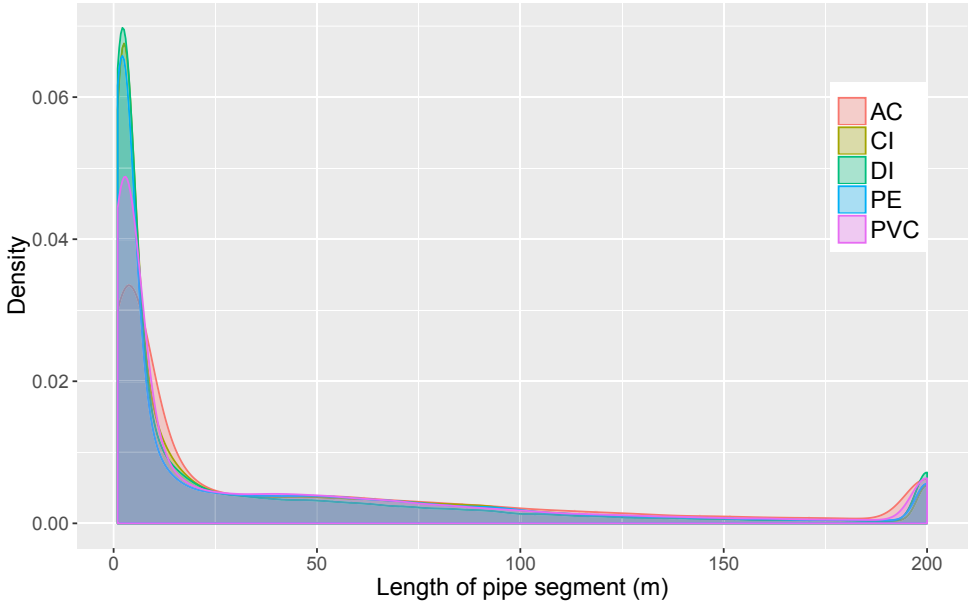


Figure B.1.: Kernel density of pipe segment length for various pipe material

B.4. Grid development

One method to mitigate errors from pipe segment representation is development of a grid system, overlaying the pipe network, where each grid unit is associated to a pipe segment. Predictor variables are assigned to each grid, by employing spatial analysis methods. For the outcome variable (pipe failure), representing 0 or 1, as whether a pipe failure event has occurred in the grid. Each grid forms an observation record in the sample set of the model.

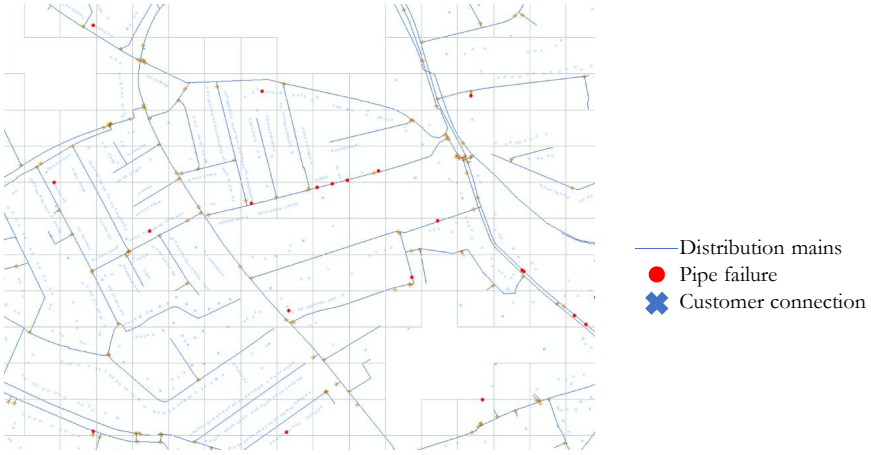


Figure B.2.: AR CG 50m grid with mains, pipe failure records and customer connections

A sampling grid is developed (square grid of 50m), where each grid represents an sample shown in Figure B.2, illustrating a section of the grid, with associated failure points, valves and customer connections in each cell. Each cell is an observation in the dataset used to model development. Predictor variables are inherited from the pipe in the grid providing the characteristics of each grid cell (e.g. pipe diameter, material, etc.). Where a grid contains more than one pipe segment (i.e. pipes with different diameter) or grid cell with more than one type of soil category, the dominant characteristic is assigned to the grid cell.

Grid size Choice of grid size is driven by a number of factors. Any grid size smaller than approximately 25m could result in exclusion of a number of customer connections to the pipes. Customers are represented as nodes in the GIS, with distribution of their distance to the closest pipe illustrated in Figure B.3.
For logistic regression model, section 8.1, selecting smaller grids will result in larger number of constraints in likelihood function (an over-determined

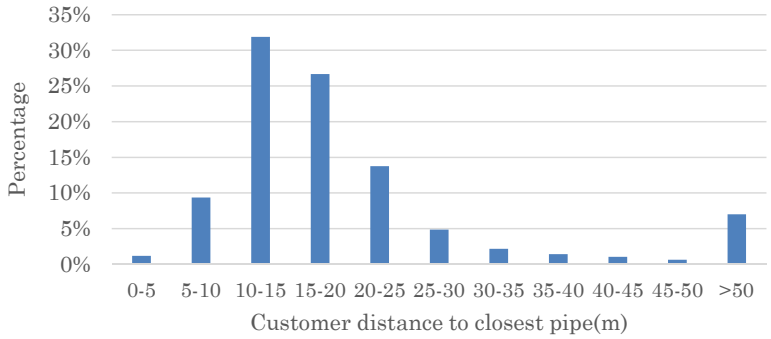


Figure B.3.: Distribution of customers distance to closest pipe

system) while the degrees of freedom would stay the same.
Selecting very large grid size would result in each grid including potentially a number of pipes and therefore errors in associating other predictors, such as pipe diameter and material.
Considering these constraints, a grid size of 50m is selected. A relatively small percentage of customers have more than 50m distance to their closest pipes. This is expected for some properties, mostly in the rural areas and industrial consumers with long supply pipes. For these cases, properties are assigned to the closest grid , even though the grid does not contain them following spatial analysis.

B.5. Correlation tables and plots

Below are the correlation tables for the models developed in chapter 8. The tables represent the correlation between various covariants considered for the model development, with bold font indicating correlation coefficients larger than 0.5.

Table B.3.: Correlation matrix - Model 1

	age	diameter	length	Pmean	propertyNo	valveNo	DP
age	**1.000**	-0.051	-0.029	0.044	0.082	0.016	-0.061
diameter	-0.051	**1.000**	0.300	0.329	-0.216	-0.021	-0.061
length	-0.029	0.300	**1.000**	0.075	0.356	-0.013	0.010
Pmean	0.044	0.329	0.075	**1.000**	-0.125	0.029	-0.011
propertyNo	0.082	-0.216	0.356	-0.125	**1.000**	-0.010	0.049
valveNo	0.016	-0.021	-0.013	0.029	-0.010	**1.000**	-0.014
DP	-0.061	-0.061	0.010	-0.011	0.049	-0.014	**1.000**

Table B.4.: Correlation matrix - Model 3

	age	diameter	length	propertyNo	valveNo	Pmean	DP
age	**1.000**	-0.099	-0.032	0.021	-0.024	-0.122	-0.010
diameter	-0.099	**1.000**	0.205	-0.182	0.001	0.217	-0.017
length	-0.032	0.205	**1.000**	0.215	0.009	0.165	-0.021
propertyNo	0.021	-0.182	0.215	**1.000**	-0.026	-0.130	0.012
valveNo	-0.024	0.001	0.009	-0.026	**1.000**	0.031	0.004
Pmean	-0.122	0.217	0.165	-0.130	0.031	**1.000**	-0.092
DP	-0.010	-0.017	-0.021	0.012	0.004	-0.092	**1.000**

Table B.5.: Correlation matrix - Model 4

	length	diameter	propertyNo	valveNo	age	Pmean	DP
age	0.006	-0.090	0.026	-0.012	**1.000**	-0.028	0.030
diameter	0.069	**1.000**	-0.130	-0.021	-0.090	0.047	-0.015
length	**1.000**	0.069	0.276	0.000	0.006	0.135	0.066
propertyNo	0.276	-0.130	**1.000**	-0.041	0.026	-0.084	-0.015
valveNo	0.000	-0.021	-0.041	**1.000**	-0.012	0.023	0.007
Pmean	0.135	0.047	-0.084	0.023	-0.028	**1.000**	0.254
DP	0.066	-0.015	-0.015	0.007	0.030	0.254	**1.000**

Table B.6.: Correlation matrix - Model 2

	age	diameter	valveNo	propertyNo	FP	corrosion	Pmean	DP
age	**1.000**	-0.239	0.009	0.164	-0.110	-0.032	-0.060	-0.080
diameter	-0.239	**1.000**	-0.026	-0.292	0.099	0.117	0.497	-0.045
valveNo	0.009	-0.026	**1.000**	-0.068	0.000	-0.004	-0.005	-0.007
propertyNo	0.164	-0.292	-0.068	**1.000**	-0.046	-0.069	-0.154	-0.028
FP	-0.110	0.099	0.000	-0.046	**1.000**	**0.501**	-0.107	-0.023
corrosion	-0.032	0.117	-0.004	-0.069	**0.501**	**1.000**	0.065	-0.036
Pmean	-0.060	0.497	-0.005	-0.154	-0.107	0.065	**1.000**	-0.069
DP	-0.080	-0.045	-0.007	-0.028	-0.023	-0.036	-0.069	**1.000**

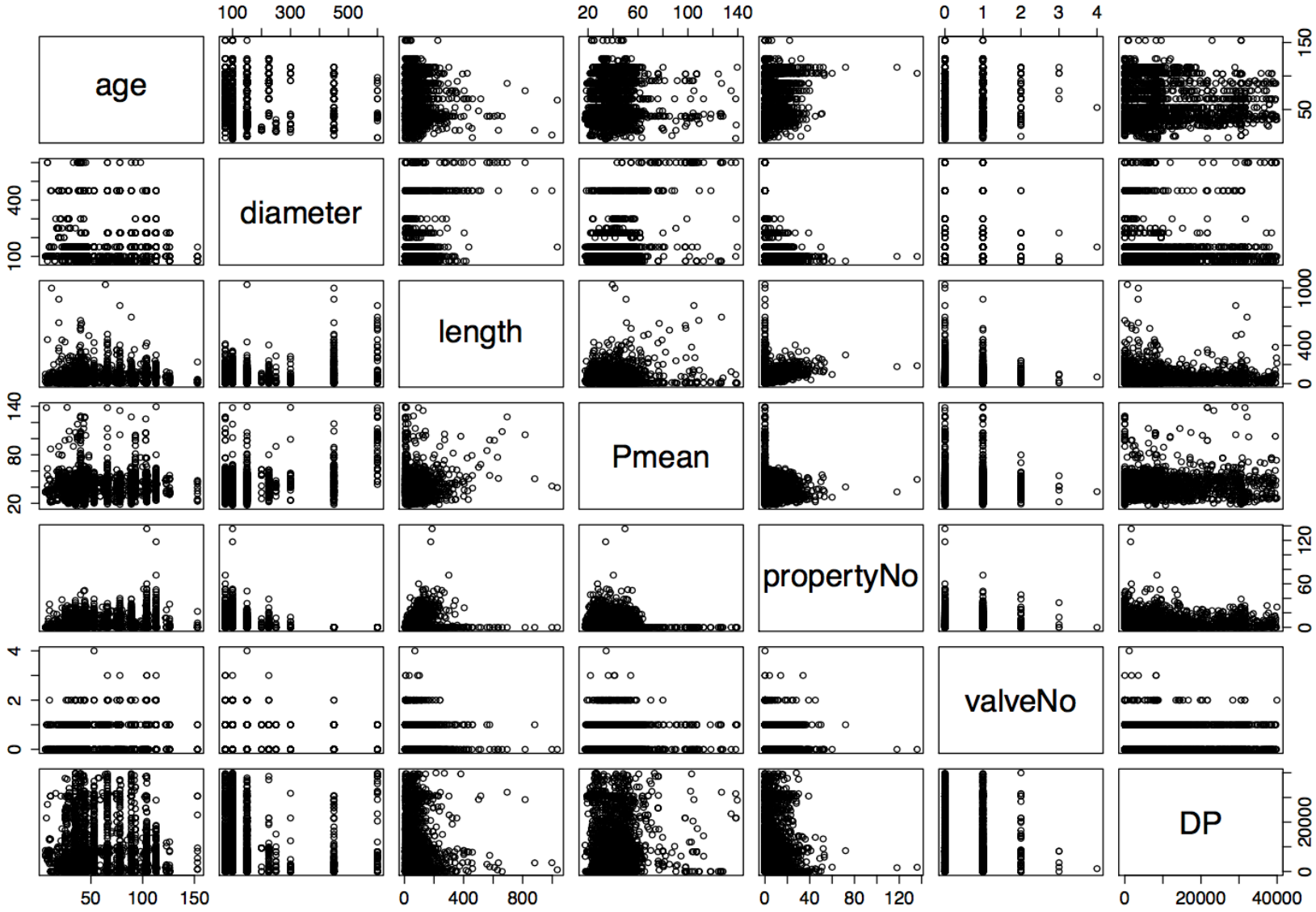


Figure B.4.: Correlation plots - Model 1

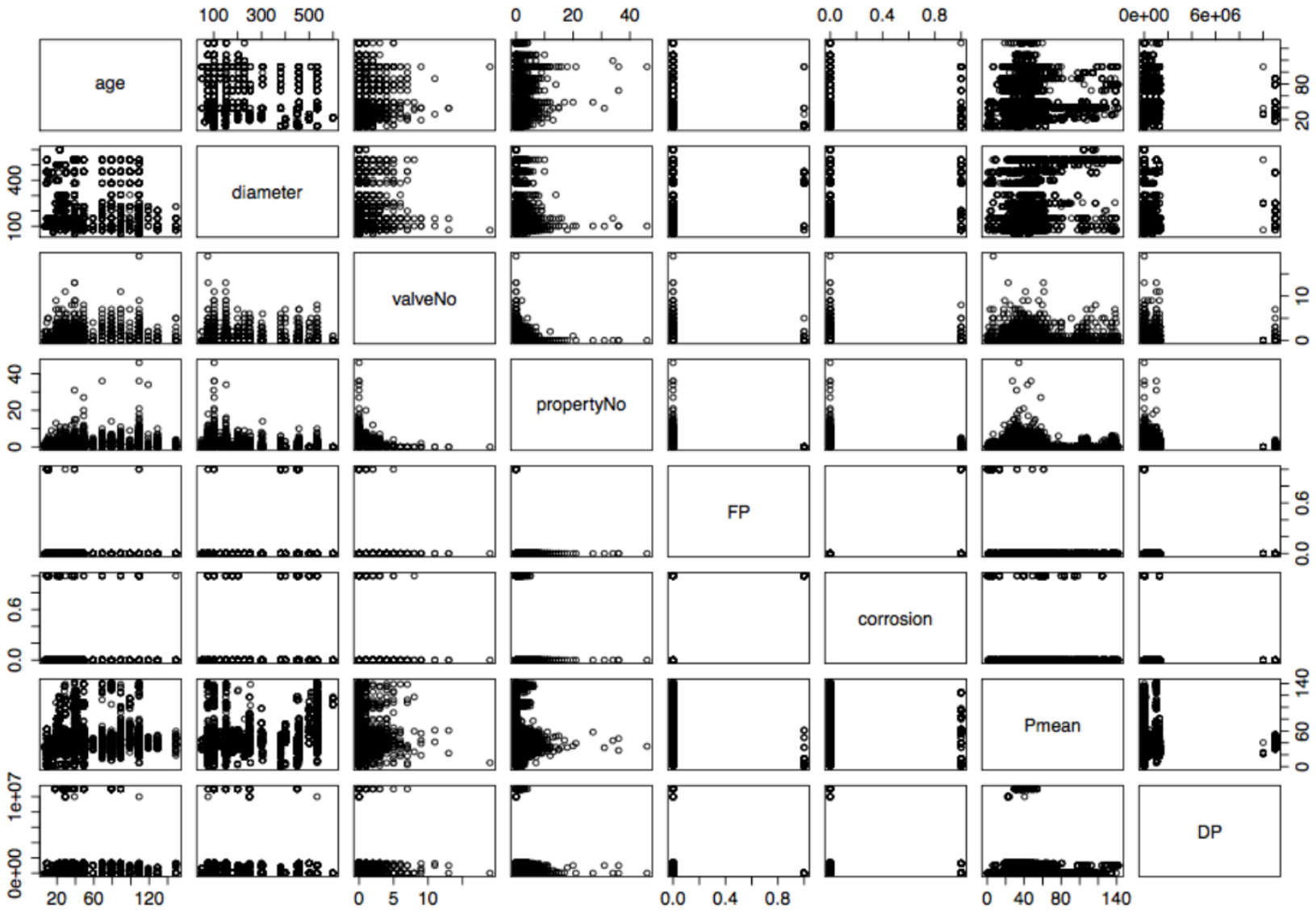


Figure B.5.: Correlation plots - Model 2

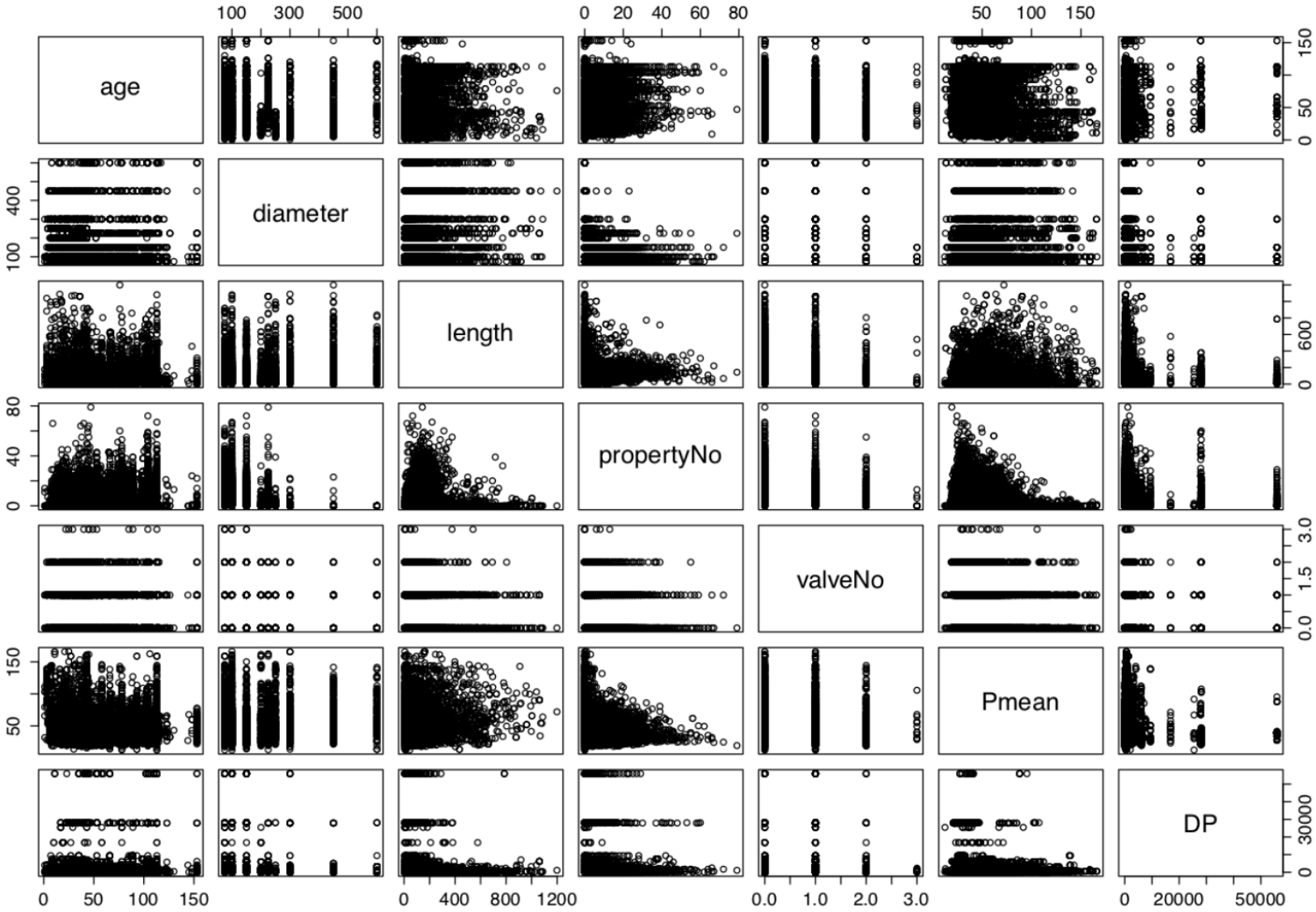
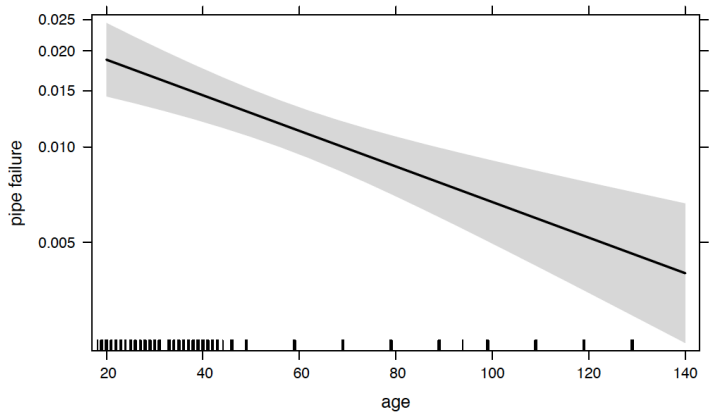
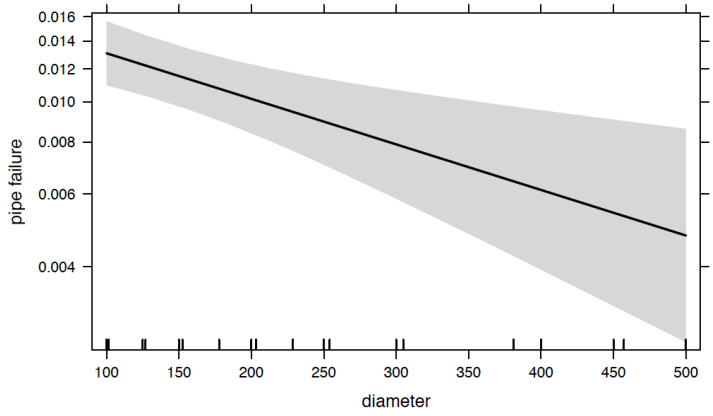


Figure B.6.: Correlation plots - Model 3

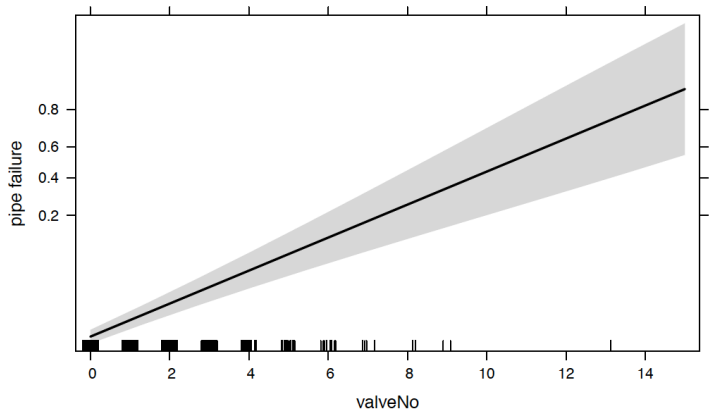
B.6. Statistical model 2 effect plots



(a) Model 2 effect plot - age

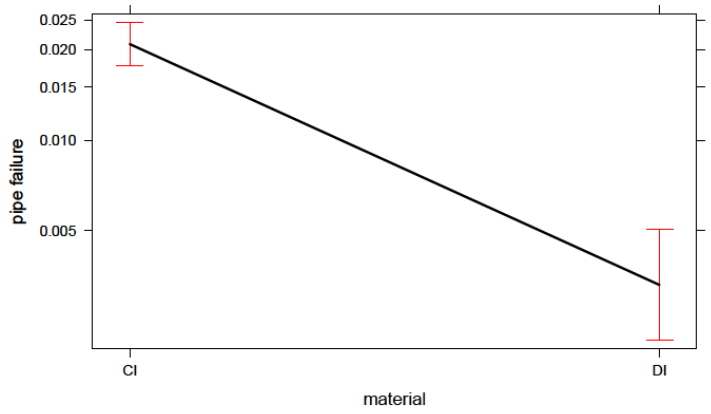


(b) Model 2 effect plot - diameter

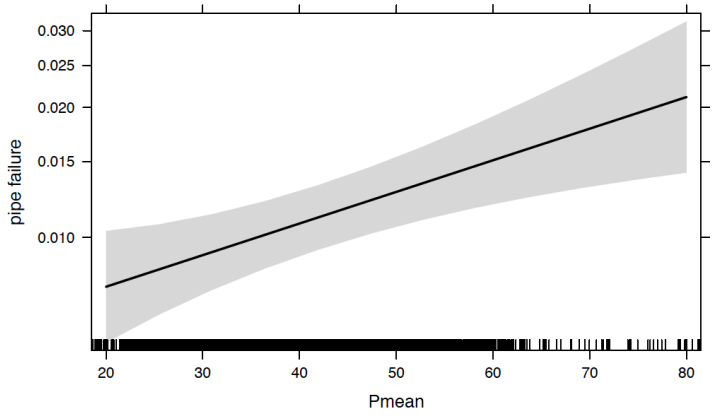


(c) Model 2 effect plot - length

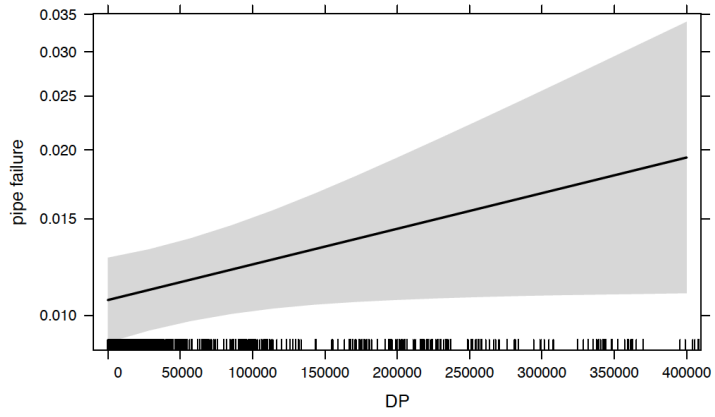
Figure B.7.: Effect plot for model 2



(a) Model 2 effect plot - material



(b) Model 2 effect plot - Pmean



(c) Model 2 effect plot - DP

Figure B.8.: Effect plot for model 2-cont.

C. Cost-benefit framework

C.1. Pipe failure cost questionnaire

Total Cost of Pipe Failures

If actual cost is not available, what is your estimate of cost in such an event

Place costs only once, please do not double count.

1	**Description of the Water Main Failure**	
	Date Main Failed	dd/mm/yyyy
	Location- X , Y Coordinate	
	Diameter of Main	mm OR inches
	Water Main Material	CI, DI, Steel, PVC, PE, AC, or other
	Root Cause of Failure	
	Burst Detection Active (Leakage Control (ALC), Visible Leak (VL), Sudden Burst (SB))	
	Type of Failure (Longitudinal Split (LS). Cracked Bell (CB), Broken Wire/ Cylinder (BW C.PCCP), Circumferential Ring Crack (RC), Joint Leak (JL), Hole in Pipe (HP) or Specify other)	

2	**Labour Costs (including all overheads)**	**Actual (£)**	**Minimum Estimated Cost (£)**	**Maximum Estimated Cost (£)**
	Construction/ Field Service/ Distribution/ Engineering Staff Labour Cost			
	Customer Service Staff Labour Cost			
	Supply interruption/ Finance/ Public Relations/ Laboratory/ Natural Resources/ IT Staff Labour Cost			
	Other Labour Costs - please list below			

3	**Purchases for the Failure Event**	**Actual (£)**	**Minimum Estimated Cost (£)**	**Maximum Estimated Cost (£)**
	Cost of Pipe/valve Purchases for the Event or Taken From Stock			
	Other Equipment Purchased for the Event (Dewatering/ Lighting and Generation/ Line Stopper/ etc.)			
	Road Repair Material Purchases for the Event			
	Total Equipment Rental Costs			
	Other Purchases for the Failure Event - please list below			

4	**Costs of Outside Service**	**Actual (£)**	**Minimum Estimated Cost (£)**	**Maximum Estimated Cost (£)**
	Total Costs for Outside Construction Contractors			
	Engineering Consulting Costs			
	Costs for Landscaping/Restoration Services/Outside Laboratory Services			
	Attorney Fees and Other Legal Costs			

Figure C.1.: Pipe failure questionnaire - page 1

Costs for Police Assistance/Fire Department Assistance/Assistance from other Water Utilities			
Costs for Utility Damages and Services (Electric/ Telephone/ Gas/ Wastewater/Transportation/ Parking/ etc.)			
Other Outside Services not listed above (including Assistance from other Water Utilities) - please list below			

5 **Regulatory Costs**	**Actual**	**Minimum Estimated**	**Maximum Estimated**
Duration of Customer Water Outage (Hours)			
Property-Minute of Residential households without water			
Property-Minute of commercial and industrial without water			
Number of critical customers (eg hospitals) or large industrial users without water			
ODI - Total supply interruption cost (£)			
Number of discolouration contacts (# of contacts)			
ODI - Total cost of discolouration contacts (£)			
Total cost of impact on SIM Score (£)			
Other Costs not included above - please list below			

6 **Other Costs**	**Actual**	**Minimum Estimated**	**Maximum Estimated**
Number and Total cost of residential dwellings damaged by flooding (£)			
Number and Cost of vehicles, commercial and industrial properties damaged by flooding (£)			
Claims Paid by the company or the Insurance (not included above)			
Cost of Water Lost			
Cost to Provide Alternative Supplies			
Customer Notification Costs			
Costs Associated with Boil Water Notice/ Test Pits/ Providing Tankers/ Providing Bottled Water/ etc.			
Other Costs not included above - please list below			

7 **Societal Cost Impact**	**Actual**	**Minimum Estimated**	**Maximum Estimated**
Estimated cost impact to the Road Systems/ Parking Authorities/ other Public Transportation Systems			
Estimated cost impact on traffic			
Estimated cost impact on local police/ fire services/ area hospital/emergency medical agencies/ etc.			
Estimated population notified of boil water notice			
Estimated number of illnesses related to the failure event			
Estimated cost impact on other utilities (Electric/ Telephone/ Gas/ Wastewater/Transportation/ Parking/ etc.)			
Other costs-impact not included above - please list below			

Figure C.2.: Pipe failure questionnaire - page 2