Imperial College of Science, Technology and Medicine
Department of Civil and Environmental Engineering

Methods for evaluating the impact of pressure transients on water distribution pipe breaks

Carlos Ingnacio Jara Arriagada

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy of Imperial College London and
the Diploma of Imperial College London, January 2025

Statement of originality

I declare that this thesis was composed by myself and the findings presented in this thesis are the product of my own research work. Any information derived from the literature or material that is not original to this work has been duly acknowledged in the thesis and the bibliography.

Carlos Ingnacio Jara Arriagada

Copyright declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Acknowledgements

This work was funded by the National Agency for Research and Development (ANID)/ Scholarship Program/DOCTORADO BECAS CHILE/2020–72210314. Their financial support is gratefully acknowledged.

I dedicate this thesis to my parents, Yanet and Carlos who have always been there for me with unconditional support, and to my brother Nicolás. Despite the distance and the little talk, I was always sure of the support you were giving to me.

I greatly thank my supervisor, Dr Ivan Stoianov for all his encouragement and support over the course of this research. His passion for the field of water distribution networks certainly impacted in my motivation and the quality of my research.

Pursuing a PhD is definitely a fantastic, but also a daunting task, and I could not have made it through without the many interesting people I have met. I particularly thank all the former residents and the office of William Temple House. I also thank the people from 409, both past and present, and the many other fellow colleagues in the Department of Civil and Environmental Engineering who made my days at Imperial more enjoyable. I am also grateful to Amigos UK; despite not meeting often, I appreciate their friendship. Finally, thank you, Diana, for all your support during the last stage of my PhD.

Abstract

Water distribution networks worldwide are deteriorating, posing significant challenges for maintenance and repair due to their vast extent and the increasing frequency of pipe breaks. Effective management of these networks requires understanding the factors driving pipe breaks. Among these factors, water pressure emerges as a critical variable that can be regulated to reduce pipe deterioration. The concept of cumulative pressure-induced stress (CPIS) was recently introduced to capture the combined impacts of both mean pressure and pressure dynamics on pipe integrity. However, understanding of water pressure as a cause of pipe breaks and the development and testing of CPIS remain areas needing further investigation.

This thesis advances the understanding of CPIS through three key components: motivation and theoretical foundations, interpretation and analysis, and practical application to operational networks. First, the impact of mean pressure and pressure range metrics on pipe breaks is analysed using a large dataset, providing empirical evidence for the relationships of quasi-steady state pressures and pipe breaks. This analysis motivated the development of a fracture mechanics-based theoretical study to explain how pressure dynamics influence pipe deterioration, forming the foundation for CPIS. Second, the predictive capability of CPIS on pipe breaks is examined using machine learning models, which demonstrate its potential to enhance pipe break predictions and identify areas in need of pressure management. Finally, the practical challenges of implementing CPIS in operational networks are addressed. This includes investigating sampling resolution requirements, evaluating an interpolation method for CPIS, and formulating and testing a transient source localisation method to guide CPIS interpolation.

The insights and methods presented in this thesis enable the application of CPIS in both large-scale experimental and operational settings. These contributions will improve the resilience and efficiency of water distribution systems, supporting the development of sustainable and reliable water distribution networks.

Contents

1 Introduction **29**

1.1 Background and motivation 29

1.2 Project aim and objectives 33

1.3 Research contributions and novelties 34

1.4 Thesis outline 40

1.5 Research dissemination 43

1.5.1 Peer-reviewed journals 44

1.5.2 Conference papers 44

2 Literature review **46**

2.1 Pipe break analysis in the water industry 46

2.2 Water pressure and pipe breaks 50

2.3 Pressure fluctuations and fatigue failure of pipes 51

2.4 The cumulative pressure induced stress 54

2.5 High frequency pressure monitoring 58

2.6 Technological challenges for CPIS interpolation 59

2.7 Localisation of sources of pressure transients 61

2.8 Summary 63

3 Average water pressure metrics and pipe breaks 66

3.1 Introduction 66

3.2 Methods 68

3.2.1 Data preparation 68

3.2.2 Logistic regression 69

3.2.3 Model fitting 71

3.2.4 Sensitivity analysis 75

3.3 Case study 76

3.3.1 Dataset summary 76

3.3.2 Exploratory analysis 76

3.3.3 Cohorts for analysis and model building 81

3.4 Results and discussion 82

3.4.1 Main model results 82

3.4.2 Sensitivity analysis 86

3.5 Conclusions 90

4 Fracture mechanics of pressure-induced fatigue failures 93

4.1 Introduction 93

4.2 Cast iron pipes in water distribution networks 94

4.2.1 Manufacturing processes of cast iron and the effect on pipe wall thickness . . . 95

4.2.2 Microstructure of cast iron pipes . . . 96

4.2.3 Graphite flakes as crack initiators . . . 97

4.2.4 The effects of corrosion on pipe wall thickness, pits and cracks . . . 98

4.2.5 Fatigue experiments on cast iron pipes . . . 101

4.2.6 Observed failures in cast iron pipes . . . 104

4.3 Assessment of water pressure fluctuations on fatigue failures in cast iron pipes . . . 110

4.3.1 Loading conditions for water distribution pipes . . . 110

4.3.2 Distributions of size and pipe wall thickness for cast iron pipes . . . 111

4.3.3 Stress variations due to a reduction in pipe wall thickness . . . 112

4.3.4 Fracture mechanics analysis . . . 112

4.3.5 Fatigue crack propagation . . . 114

4.4 Results from fracture mechanics evaluation and fatigue crack propagation . 118

4.4.1 Effect of the reduction in pipe wall thickness on pipe stresses . . . 118

4.4.2 Maximum crack lengths in cast iron pipes . . . 119

4.4.3 Minimum magnitude of water pressure cycles for pressure-induced fatigue . . . 121

4.4.4 Fatigue crack propagation . . . 122

4.5 Discussion . . . 127

4.6 Conclusion . . . 131

5 Effect of CPIS on pipe break prediction models 133

5.1 Introduction 133

5.2 Methodology 135

5.2.1 Data curation and exploratory analysis 136

5.2.2 Random forest regression and variable performance 136

5.3 Case study – Description of the networks 138

5.3.1 NetA description 139

5.3.2 NetB description 139

5.4 Results and discussion 141

5.4.1 NetA 141

5.4.2 Development of models for NetB 146

5.4.3 Assessment of the impact of the cumulative pressure-induced stress as a metric in pipe break models 148

5.5 Conclusion 150

6 High resolution pressure monitoring for CPIS estimation 151

6.1 Introduction 151

6.2 Methods 152

6.2.1 Pressure monitoring technologies 152

6.2.2 Cycle counting 153

6.2.3 Assessment of the required sampling rate 154

6.2.4 Dynamic pressure profile assessment 155

6.3 Results . . . 156

6.3.1 Sampling rate . . . 156

6.3.2 Dynamic vs calm profiles . . . 158

6.4 Conclusion . . . 160

7 Interpolation of CPIS across water networks 162

7.1 Introduction . . . 162

7.2 Methodology . . . 163

7.2.1 The DP variable . . . 164

7.2.2 Temporal interval for DP calculation . . . 165

7.2.3 Threshold for DP separation . . . 166

7.2.4 DP above threshold interpolation, DP_{high} . . . 168

7.2.5 DP below threshold interpolation, DP_{low} . . . 168

7.2.6 Spatial interpolation methods . . . 169

7.2.7 Inverse distance weighting interpolation . . . 170

7.2.8 Graph Laplacian-based interpolation . . . 171

7.2.9 Leave-one-out cross-validation . . . 173

7.2.10 Transient model calibration . . . 173

7.3 Case study . . . 174

7.3.1 Transient pressure profiles . . . 175

7.4 Results . . . 178

7.4.1 Determination of the DP separation threshold . . . 178

7.4.2 Variograms for DP variables 181

7.4.3 Interpolation results 181

7.5 Discussion 186

7.6 Conclusion 188

8 Localisation of sources of pressure transients 190

8.1 Introduction 190

8.2 Methodology 192

8.2.1 Problem formulation 192

8.2.2 Solution method 199

8.2.3 Measurement of errors in results 202

8.2.4 Experimental setup 202

8.3 Numerical experiments 203

8.3.1 Simulated case studies 204

8.4 Field case study validation 210

8.4.1 Case study overview 210

8.4.2 Pressure transient onset detection 212

8.4.3 Evaluation trials and estimation of errors in the solution 213

8.4.4 Influence of the number of sensors 214

8.5 Results 214

8.5.1 Scenario 1 214

8.5.2 Scenario 2 217

8.5.3 Scenario 3 219

8.5.4 Additional testing scenarios 220

8.5.5 Field case study validation results 223

8.6 Conclusions 227

9 Conclusion **231**

9.1 Summary of thesis achievements 231

9.1.1 Quantification of the impact of steady-state mean pressure and pressure range with pipe breaks 232

9.1.2 Fracture-mechanics-based analysis of pressure fluctuation magnitudes and mean pressure levels that lead to fatigue failure in deteriorated pipes 233

9.1.3 Evaluation of the use of CPIS in machine learning models for pipe break prediction 234

9.1.4 Determination of the optimal sampling rate for reliable cycle counting in water pressure time series used in CPIS calculation 236

9.1.5 Method for interpolating a dynamic pressure metric used in CPIS calculations for each pipe in a network 237

9.1.6 Method to localise sources of transients in networks with uncertain connectivity 238

9.2 Limitations and future work 238

Bibliography **242**

A Transient model calibration **263**

A.1 Genetic algorithm optimisation of transient parameters 264

A.2 Post-processing of transient model results 264

B Additional testing scenarios **268**

C Details of the weight function **273**

D Measured (from simulation) vs analytical pressure wave arrival times **276**

E Permissions to reproduce own published work **278**

List of tables

3.1 Case study. 77

3.2 Pipe cohorts for modelling. 81

3.3 Variables used as input for modelling. 81

3.4 Performance metrics by pipe cohorts and maximum polynomial degree explored by the stepwise regression using the Akaike Information Criterion. . 84

3.5 Covariates and coefficients for the logistic regression models utilised for analysis in every pipe cohort. 85

3.6 Effect of reducing the mean pressure in the model's predictions. 90

3.7 Effect of reducing the pressure range in the model's predictions. 90

4.1 Reported measured mechanical properties for cast iron. 98

6.1 Cycle counting comparison at different time resolutions, 1 year of data 2018.156

6.2 Cycle counting comparison at different time resolutions, 1 year of data 2021.157

7.1 Leave-one-out cross-validation evaluation for interpolation methods. 182

8.1 Pipe material properties used to determine pipe wave speeds in the network.211

8.2 Validation scenarios for the solution of Problem (8.7). 222

D.1 Comparison measured (from transient simulation) vs analytical pressure wave arrival times for scenario 1. 277

List of figures

1.1 Schematic of failure rates in the life of cast iron pipes (bathtub curve). . . 30

3.1 Central zone of the water distribution network. 78

3.2 Distribution of pressure metrics for CI pipes. 79

3.3 Number of breaks by pipe length range. 79

3.4 Seasonal break rates. 80

3.5 Areas under curve (AUC) of the receiver operating characteristics (ROC) and precision-recall (PR) curves for training and cross-validation (cv) by increasing model complexity. 82

3.6 Coherence check for the asbestos cement model up to quadratic terms by varying values of age, length and diameter for all pipes. 83

3.7 Pipe break probability change by decreasing pressure metrics. 87

4.1 Variability and non-uniformity in pipe wall thickness due to manufacturing practices. 96

4.2 The extent of graphitic corrosion in a cast iron pipe as identified by sand-blasting. 99

4.3 Pipe wall thicknesses for pit and spun cast iron pipes according to British Standards BS-78 and BS-1211, and comparison to maximum observed corrosion pit depths. 101

4.4 Fatigue crack propagation curves for cast iron at different loading ratios R and sources. 102

4.5 Example of an exhumed pipe provided by the water utility. 106

4.6 Examples of pipe deterioration due to corrosion pits and patches after sandblasting pipes. 107

4.7 Longitudinal failures in water distribution pipes due to initial corrosion pits. 108

4.8 Examples of cast iron pipe blowouts. 109

4.9 Pipe blowout with the presence of longitudinal crack. 109

4.10 Pressure transient in a water distribution pipe acquired from an operational network. 111

4.11 Illustration of the empirical relation between the rate of crack propagation (da/dN) and stress intensity factor range (ΔK). 115

4.12 Walker equation fit for a selected representative cast iron material. 117

4.13 Effect of pipe wall thickness reduction on the D/2t ratio for different cast iron pipe diameters. 119

4.14 Maximum crack length for cast iron pipes as a function of increasing water pressure and pipe wall thickness reduction. 120

4.15 Minimum water pressure to reach stress intensity threshold for cast iron pipes. 122

4.16 Effect of pipe wall thickness on the number of cycles to failure for different initial crack lengths and water pressure fluctuations. 123

4.17 Fatigue life estimation with a reduction in pipe wall thickness at varied cyclic pressure loadings and mean pressure. 124

4.18 Fatigue life estimations for varying magnitudes of pressure fluctuations at constant magnitudes of mean pressure. 125

4.19 Comparison cycle range reduction versus mean pressure reduction. 126

4.20 Effect of mean pressure reduction and increase of pressure cycles. 127

5.1 The approach followed in the two WDSs. 135

5.2 Classification of network characteristics for NetA and NetB. 140

5.3 NetA results. 143

5.4 Variable importance estimates and interaction strength for NetA. 144

5.5 Variable importance estimates and interaction strength for NetB. 148

6.1 Installation of pressure transducers. 152

6.2 Comparison of pressure profiles obtained by a low and high frequency sampling rates. 153

6.3 Visualisation of cycle detection by the rainflow cycle counting technique. . 154

6.4 Measured water pressure profiles in operational water distribution networks, example of a regular day. 159

6.5 Cycle count histogram comparison of dynamic vs calm profiles over one year of pressure monitoring. 159

7.1 Flowchart for threshold determination. 167

7.2 Case study network and sensor's location. 175

7.3 Example high-resolution pressure data downsampled at 1 sample per second. 176

7.4 DP magnitude variation with increasing minimum cycle range. Dataset 1. . 177

7.5 Estimation of DP separation threshold. 179

7.6 DP separation after 4m threshold setting. 180

7.7 Daily DP variations. 181

7.8 Variograms for DP variables. 182

7.9 DP_{high} interpolation. 184

7.10 DP_{low} interpolation. 186

8.1 Flowchart for new transient source localisation. 201

8.2 Network net3 used for testing the transient source localisation algorithm. . 207

8.3 Time series of pressure transient for the presented scenarios. 208

8.4 Network model and sensors' locations. 211

8.5 Detection of dropping transient pressure onset arrival time using the gradient method. 213

8.6 Solutions for the scenario 1 for transient source localisation by different algorithms. 216

8.7 Solution for the scenario 2 for transient source localisation. 218

8.8 Solution for scenario 3 for transient source localisation. 221

8.9 Results of the pressure transient source localisation method, on a transient event recorded near 03:00 AM 223

8.10 Distribution of pressure wave travel times with reference from Sensor 1. . . 224

8.11 Distribution of source localisation errors. 226

8.12 Influence of number of sensors in source localisation errors. 226

A.1 Transient pressure simulation and post-processing of transient pressure Sensor 107 266

A.2 Comparison adjusted transient model against actual measured data for all sensor readings in dataset 1 267

A.3 Comparison threshold estimation using original and adjusted transient model results. 267

B.1 Scenario 4. 269

B.2 Scenario 5 269

B.3 Scenario 6 270

B.4 Scenario 7 270

B.5 Scenario 8 271

B.6 Scenario 9 271

B.7 Additional testing scenarios. 272

C.1 Example of the different paths to be analysed by the algorithm. 275

C.2 The purpose of the weight function on additional time added to analytical times. 275

Acronyms

The following list consists of a non-exhaustive compilation of acronyms used throughout the thesis. Other specific acronyms, symbols and definitions used in individual chapters are presented in their corresponding methodological sections.

AC Asbestos Cement

AIC Akaike Information Criterion

AZP Average Zone Pressure

BIC Bayesian Information Criterion

CI Cast Iron

CPIS Cumulative Pressure-Induced Stress

DMA District Metered Area

DP Dynamic Pressure

GIS Geographic Information System

GL Graph Laplacian

IDW Inverse Distance Weighting

LOOCV Leave-One Out Cross Validation

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

PRV Pressure Reducing Valve

RF Random Forest

RMSE Root Mean Squared Error

SCADA Supervisory Control and Data Acquisition System

WDN(s) Water Distribution Network(s)

Units and measurement system

All measurements in this thesis follow the metric system. Otherwise stated, all water pressure values in this thesis are expressed in meters of water column (m or mH_2O), which is the industry standard unit for water pressure in water distribution networks. The conversion to the SI unit of pressure, Pascal (Pa), follows the relation: 1 mH_2O = 9,806.65 Pa.

Chapter 1

Introduction

This thesis investigates the evidence, theoretical foundations, and methods for evaluating the concept of cumulative pressure-induced stress (CPIS) within water distribution networks. CPIS is an innovative concept developed to characterise pressure dynamics within water pipes, and aims primarily at associating the impact of cumulative stresses on pipe breaks. A wide range of interdisciplinary methods are used for the purposes of this thesis, encompassing topics of statistics, machine learning, fracture mechanics, signal processing, optimisation and pipe hydraulics. This chapter describes the background and motivation of this investigation, presents the research objectives, highlights the research contributions, and outlines the contents of the following chapters.

1.1 Background and motivation

The water industry's principal objective is to provide safe water, sanitation, and drainage services. These core services underpin modern society, public health and the efficient operation of numerous other sectors. Consequently, reliable and properly maintained water infrastructure is indispensable for human development, public health and economic progress.

Water distribution systems in many parts of the world incorporate pipeline infrastructure

that have been in continuous operation for over a century. For example, in the UK and the USA, a significant portion of their water distribution networks (WDNs) consists of cast iron pipes installed during the Victorian era (Marlow, Gould, Beale & Lane, 2015; Marlow, Gould & Lane, 2015). At that time, design standards were still in their infancy, leading to an oversizing approach for civil infrastructure, including water distribution pipes. This approach led water distribution pipes to remain in service for much longer times than their planned design life.

Many of these oversized cast iron pipes have remained operational into the present day. In the UK, cast iron pipes constitute an estimated 30% to 60% of the WDNs in several urban areas (Thames Water, 2023). However, these ageing pipelines are increasingly susceptible to corrosion and added stresses from climate change and urban development. As a result, many of these pipes are now entering the wear-out phase of their service life (Figure 1.1).

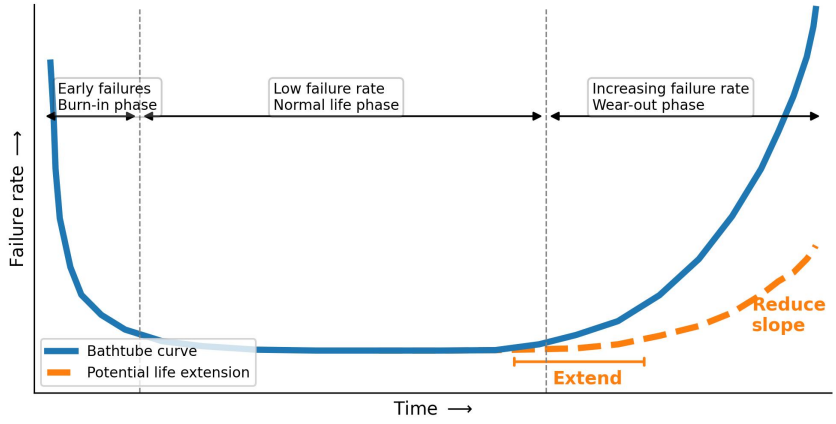


Figure 1.1: Schematic of failure rates in the life of cast iron pipes (bathtub curve). From: Jara-Arriagada & Stoianov (2024).

The wear-out phase indicates that the components of a system have entered a stage of accelerated deterioration reaching the end of their functional lifespan and becoming increasingly vulnerable to normal loading conditions. Therefore, these components should be repaired or replaced. However, given the extent of existing pipelines and the industry's current replacement and rehabilitation rates, ranging from 0.8% to as low as 0.1% per year, Victorian-era pipes may remain in service for several more decades (Economic Insight, 2022). This problem is not limited to cast iron, and other pipe materials are

similarly affected. If no proactive actions are taken, the water industry could encounter an unprecedented escalation in pipe breaks.

Currently, according to Barfuss (2023), the US and Canada experience a combined total of 260,000 water main breaks each year, representing approximately $2.6 billion in annual maintenance and repair costs. Furthermore, over the next 20 years, an estimated $420.8 billion is expected to be needed for the maintenance of underground distribution and transmission pipes in the US, with similar proportional costs anticipated for Canada. This represents 60% of the total funding required by the water industry (Environmental Protection Agency, 2023). The economic impact of pipe breaks, the need for effective future maintenance strategies, and the social consequences related to health, safety, and customer satisfaction have driven significant research into predicting pipe breaks and developing strategies to mitigate their causes.

Various factors contribute to pipe deterioration and failure, and are commonly classified into three categories (Barton et al., 2019): (1) environmental factors, such as soil corrosion, ground movement, and freeze-thaw cycles; (2) intrinsic factors related to the pipe itself and its installation, including diameter, material, joints and backfill; and (3) operational factors, primarily involving water pressure metrics and their fluctuations. Among these, water pressure is particularly significant due to the potential for its magnitude to be controlled using advanced pressure management techniques (Changklom & Stoianov, 2019; Martin et al., 2022).

The interest in investigating water pressure has been further supported by increasing evidence of frequent pressure fluctuations within water networks (Starczewska et al., 2014; Rathnayaka et al., 2015). These fluctuations impose cyclic loadings on pipes, which can lead to fatigue failure. Additionally, numerous studies have demonstrated a correlation between steady-state water pressure and pipe breaks (Martínez-Codina et al., 2016; Martínez García et al., 2020; Moslehi & Jalili_Ghazizadeh, 2020). The hypothesis that water pressure is linked to pipe breaks is promising, as technologies such as pressure-reducing valves and transient control devices could be employed to reduce both average zone pres-

sure (AZP) and pressure fluctuations (Changklom & Stoianov, 2019; Martin et al., 2022). By doing so, these technologies could potentially extend the lifespan of pipes, delaying the wear-out phase and providing additional time for maintenance planning and budget allocation (Figure 1.1).

Despite the potential importance of water pressure on pipe breaks in the water industry, its impact on pipe breaks has not been conclusively proven or validated. As previously stated, several studies have established statistical links between average water pressure metrics and pipe breaks, however, the effect of regulating these metrics on reducing pipe breaks remains unclear. Other studies, particularly, research dissertations by Hoskins (2015), Rezaei (2017), and Starczewska (2016) have focused on observational analysis to gather evidence of hydraulic pressure instabilities and their impact on pipe breaks in water networks. In particular, Hoskins (2015) and Rezaei (2017)'s work introduced the concept of cumulative pressure-induced stress (CPIS).

CPIS is an innovative concept inspired by fatigue analysis principles to characterise the pressure dynamics within WDNs, combining the effects of average daily mean pressure and pressure fluctuations measured at high resolution (Stoianov & Hoskins, 2017). The motivation for developing and applying the CPIS concept lied in the need to gain better understanding into how pressure dynamics lead to pipe deterioration and failures in water distribution systems (Stoianov & Hoskins, 2017). Traditional metrics often focus on static pressure measurements or average values, which fail to capture the full range of absolute pressures and the cumulative effects of fluctuating pressures on system components such as pipes, valves, and joints. These cumulative effects can lead to long-term stress accumulation, material fatigue, and ultimately, system failures or bursts (Fatemi & Vangt, 1998). By focusing on causality, the CPIS concept aims to provide insights into how operational scenarios, such as varying demand patterns, pump failures, or valve operations, contribute to stress accumulation in different parts of the network. This understanding is particularly valuable for water utilities and engineers who need to design more resilient systems, develop preventive maintenance strategies, and prioritise investments in infrastructure upgrades given the described challenges of ageing infrastructure.

Research on CPIS is, however, still in its infancy, with especial need to enhance and refine its practical application in three key areas: 1) motivations and theoretical foundations that justify its use; 2) a better interpretation framework and analysis to correlate this concept with pipe breaks; 3) and the development of methods to support its automatic calculation and application in large-scale networks. This thesis aims to address these three key areas of development, with the ultimate goal to provide the motivational and technological requirements to apply and test the CPIS in large-scale experiments. These future experiments would conclusively test the hypothesis of water pressure impact on pipe breaks, and establish if CPIS could be a critical concept and metric for managing WDNs.

1.2 Project aim and objectives

The previous section highlighted the potential of water pressure control to reduce pipe breaks, and the need for a comprehensive analysis of the CPIS in three key areas to ensure its testing as a metric for evaluating the impact of water pressure on pipe integrity. A thorough CPIS analysis requires the development of a fundamental theoretical framework and detailed models to accurately examine and characterise water pressure in WDNs. Building on the initial findings from previous research and recognising the need for more sophisticated methods and empirical evidence in CPIS analysis, this thesis aims to *refine the motivations, theoretical foundations, and methodological approaches for implementing CPIS in order to improve the asset management of water distribution networks.*

This overall aim has led to the formulation of the following objectives within the three key areas of development:

- Motivations and theoretical foundations for CPIS

    1. Review and investigate the relationships between pipe breaks and average metrics of water pressure, and the motivations for the investigation of CPIS in

water distribution networks.

2. Examine the fundamentals of fracture mechanics and material properties of cast iron pipes to inform the analysis of fatigue failure in pipes, and the conditions under which CPIS can be a relevant explanatory variable for pipe breaks.

- Interpretation and analysis of CPIS on pipe breaks

3. Investigate the impact of CPIS on machine learning models for pipe break prediction.

- Automatic computation and implementation of CPIS in large-scale networks

4. Determine the optimal pressure sampling parameters required for CPIS computation and effective energy management of battery-operated pressure monitoring devices.
5. Develop and validate a methodology to spatially interpolate the CPIS for every pipe within a water distribution network using high-resolution pressure data collected by a discrete set of monitoring devices.
6. Formulate and validate a method for localising the source of pressure transients in water networks needed for CPIS interpolation, taking into account uncertainties in network connectivity and the timing of pressure wave arrivals.

1.3 Research contributions and novelties

This thesis makes contributions to water distribution systems asset management by advancing the understanding of how pressure statics and dynamics affect pipe breaks, providing a practical framework for the interpretation and application of pressure data, and introducing tools for automated CPIS analysis. These advancements equip water utilities to strengthen their networks' resilience and efficiency, ultimately fostering more sustainable and reliable water distribution systems. In particular, the development of the formulated objectives has led to the following detailed research contributions:

1. **Quantification of the impact of steady-state mean pressure and pressure range with pipe breaks.**

   Previous research has established general relationships between steady-state water pressure and pipe breaks in WDNs using statistical tests, but they have not fully modelled how variations in water pressure influence the frequency of pipe breaks. This contribution addresses this gap by developing a non-linear logistic regression model that demonstrates how mean daily water pressure, and daily pressure range correlates with different types of pipe materials across two distinct seasons —summer and winter —and how they affect the predicted number of pipe breaks. The results show that both water pressure and pressure range significantly impact the predicted number of pipe breaks, with pressure range exhibiting a particularly strong correlation. Despite these strong correlations, the magnitudes of steady-state mean pressure and pressure range are insufficient to cause the ultimate failure of pipes. Consequently, these findings provide a foundation for hypothesising that pressure dynamics might be the underlying explanatory variable driving pipe breaks, and and thus motivate the exploration of the CPIS metric.

   The research contributed to the following publication:

   - Jara-Arriagada, C, & Stoianov, I. (2021) Pipe breaks and estimating the impact of pressure control in water supply networks. *Reliability Engineering & System Safety.* 210, 107525, ISSN: 0951-8320

2. **A theoretical review and fracture-mechanics-based analysis of pressure fluctuation magnitudes and mean pressure levels that lead to fatigue failure in deteriorated pipes.**

   The central hypothesis driving the development of the CPIS is that pressure dynamics (fluctuations) substantially contribute to fatigue failure in water distribution pipes. While previous studies have also proposed this hypothesis, it remains unsupported by fundamental theory. This contribution addresses the gap through a comprehensive review of exhumed pipe reports and a detailed theoretical fracture

mechanics analysis, focusing on the role of pressure dynamics in pipe breaks.

An extensive review of the material properties of cast iron pipes, alongside the standards for their installation in UK water networks, and an examination of the impact of corrosion, reveals that a representative ∼150 mm cast iron water distribution pipe must experience a wall loss of approximately 7 mm caused by corrosion before pressure dynamics can induce fatigue. Analysis of exhumed pipes and existing literature indicates that this condition is commonly observed in the ageing UK's water distribution networks. Consequently, this finding supports the hypothesis that CPIS can inform fatigue failure in pipes. Furthermore, this analysis establishes the critical magnitudes of water pressure fluctuations necessary for fatigue crack propagation, which are essential for the accurate computation of CPIS.

This research contributed to the following publication:

- Jara-Arriagada, C, & Stoianov, I. (2024) Pressure-induced fatigue failures in cast iron water supply pipes. *Engineering Failure Analysis*. 155, 107731, ISSN: 1350-6307

3. **Evaluation of the use of CPIS in machine learning models for improving the prediction of pipe breaks.**

   The previous contributions established a theoretical foundation for the potential of the CPIS to explain pipe breaks. While prior research applied statistical models to assess the impact of CPIS on pipe breaks, it did not investigate the role of CPIS in non-linear machine learning models. To address this gap, this contribution examines and evaluates the effect of including CPIS as an explanatory variable within Random Forest (RF) models for analysing the primary causes of pipe breaks in two WDNs. Multiple RF models—with and without pressure-related variables—were developed, and their performance, as well as the significance and interpretation of each input variable, was subsequently assessed. The results indicate that CPIS can substantially enhance pipe break prediction models, particularly in networks where pressure-related factors contribute to pipe breaks.

This research contributed to the following publication:

- Konstantinou, C., Jara-Arriagada, C, & Stoianov, I. (2024) Investigating the Impact of Cumulative Pressure-Induced Stress on Machine Learning Models for Pipe Breaks. *Water Resources Management.* 38, ISSN: 603-619, 0920-4741

4. **Determination of the optimal sampling rate for reliable cycle counting in water pressure time series used in CPIS calculation.**

   Calculation of the CPIS involves determining the number of cycles generated by pressure transients through high-resolution water pressure measurements. While higher sampling rates enhance the accuracy of capturing the rising edge of transient events, the trade-off between accuracy and sampling rate has not been thoroughly investigated. Since higher sampling rates increase energy consumption, identifying the minimal sampling rate that still provides reliable results could extend the lifespan of battery-powered pressure loggers.

   An analysis of two years of data sampled at 128 samples per second (S/s) indicates that reducing the sampling rate to 10 S/s results in an acceptable 6-11% reduction in cycle counts for pressures exceeding 5 meters. At 1 S/s, cycle counts for pressures above 5 meters decrease by approximately 15–25%. However, by applying compensatory measures to account for these missing cycles, even a sampling rate of 1 S/s may be viable for cycle counting. This adjustment opens a wider array of options for collecting high-resolution pressure data necessary for CPIS computation, while simultaneously extending the operational life of battery-powered pressure loggers.

   This research contributed to the following publication:

   - Jara-Arriagada, C, & Stoianov, I. (2022) High resolution water pressure monitoring for the assessment of fatigue damage in water distribution pipes. *Proceedings of the 2nd International Joint Conference on Water Distribution System Analysis (WDSA)& Computing and Control in the Water Industry (CCWI).* 1220-1235, ISBN: 978-84-9048-982-6

5. **Method for interpolating a dynamic pressure metric used in CPIS calculations for each pipe in a network, using time-series data captured by a discrete set of pressure loggers.**

   The cumulative pressure-induced stress (CPIS) was developed to quantify the impact of pressure fluctuations on pipe breaks. To apply this metric in operational WDNs, it is necessary to estimate its value for each pipe within the network. CPIS consists of two components: the mean daily pressure in the pipes and a "dynamic pressure" component derived from the frequency and range of pressure cycles over a specified interval. While mean pressure can be relatively easily estimated for each pipe, estimation of the number of pressure cycles for each pipe remains challenging. Long-term transient simulations are currently infeasible due to computational constraints.

   This contribution proposes an alternative method for interpolating pressure transients by integrating results from short-term transient simulations with spatial interpolation techniques commonly utilised for geospatial data interpolation.

   The findings indicate that the dynamic pressure metric can be divided into two distinct components, based on a threshold range of the cycles. Accordingly, dynamic pressure is categorised into small- and large-range contributions. The large-range component can be interpolated by scaling the outcomes of a short-term transient model, assuming that the source of these large transients is identifiable. Meanwhile, the small-range component can be interpolated using spatial interpolation methods, with the accuracy of this approach depending on the density and distribution of the available pressure loggers.

   This research contributed to the following publication:

   - Jara-Arriagada, C, & Stoianov, I. (2025) Interpolating a transient pressure metric using sensor data for water supply network asset management. Under Review. *Water Resources Research*.

6. **Method to localise sources of transients in networks with uncertain con-**

nectivity.

To accurately interpolate the CPIS for large-range transient events, it is necessary to pinpoint the sources of pressure transients. In this contribution, a novel transient source localisation method is proposed to identify transient sources even under uncertain network connectivity due to unknown valve statuses. Formulated as a mixed-integer program, the method minimises the discrepancy between analytical and measured pressure wave arrival times, leveraging the k-shortest path routing problem to account for multiple propagation routes from candidate sources to pressure loggers. The method was validated using both simulated and experimental data and additionally provide an analysis of the minimum sensor density required for successful localisation.

This research contributed to the following publications:

- Jara-Arriagada, C, Ulusoy, A.J., & Stoianov, I. (2024) Localization of Transient Pressure Sources in Water Supply Networks with Connectivity Uncertainty. *Journal of Water Resources Planning and Management.* 150(4), 107731, ISSN: 0733-9496
- Jara-Arriagada, C, & Stoianov, I. (2023) Analytical evaluation of an optimization-based transient source localization algorithm under uncertainties in water supply networks *19th Computing and Control for the Water Industry Conference (CCWI) 2023.*
- Jara-Arriagada, C, & Stoianov, I. (2024) Experimental evaluation of a pressure transient source localization method in an operational water distribution network. *15th International Conference on Hydroinformatics (HIC 2024).*
- Jara-Arriagada, C, & Stoianov, I. (2025) Field validation of a pressure transient source localization method in a water distribution network. Under Review. *Journal of Hydroinformatics.*

Overall, the results and contributions of this research dissertation have advanced the concept of the CPIS to a level where it enables the design of large-scale long-

term experimental programs to investigate the impact of pressure fluctuations on pipe breaks in water distribution networks, and guide operational strategies for extending the service life of deteriorated water networks.

1.4 Thesis outline

Chapter 2 begins by outlining the existing literature on pipe break analyses in the water sector and the primary factors contributing to pipe breaks. It then presents a comprehensive review of research on the effects of both steady-state water pressure and pressure dynamics on pipe breaks, along with the historical development and formulation of the cumulative pressure-induced stress (CPIS) concept. The chapter further addresses the technological challenges and requirements necessary for the feasibility of testing the CPIS concept in large-scale, long-term experimental programs. It concludes with a summary of the rationale behind the objectives established for this thesis.

Chapter 3 revisits the relationship between steady-state water pressure and pipe breaks by examining an extensive dataset of historical pipe repairs, correlated with mean steady-state pressure and estimated pressure range data from a large operational network. This analysis serves to confirm and evaluate the initial evidence linking water pressure to pipe breaks. The findings indicate that both mean pressure and pressure range exert a significant influence on the fitted statistical models, each showing a direct relationship with pipe breaks. A subsequent sensitivity analysis suggests that reducing these pressure metrics through pressure control approaches may substantially lower the incidence of pipe breaks. However, despite these encouraging results, the observed values of mean pressure and pressure range do not align closely with the magnitudes typically required to cause pipe breaks. This discrepancy reinforces the primary hypothesis underpinning the cumulative pressure-induced stress concept, which is that water pressure fluctuations may be the main contributing factor for the recorded pipe breaks.

Building on the conclusions of Chapter 3, Chapter 4 provides an in-depth examination of

the fundamental mechanisms by which water pressure fluctuations can induce fatigue in pipes. The analysis focuses primarily on cast iron pipes, given their extensive presence in UK water networks and their status as the oldest cohort of pipes in service. Drawing on fracture mechanics and fatigue crack propagation principles, the chapter defines the conditions under which cast iron pipes may fail due to pressure fluctuations. The findings confirm that water pressure fluctuations can significantly contribute to crack propagation in the oldest and most deteriorated segments of the network. Moreover, preventing further crack propagation can markedly extend the service life of these pipes, since even minor crack growth substantially reduces fracture toughness. Consequently, the chapter provides a theoretical validation of the potential impact of pressure fluctuations on ageing cast iron pipes, and by extension, other materials with low fracture toughness. These conclusions support the rationale for using a metric such as CPIS to explain pipe breaks in water distribution networks.

On the basis of the theoretical foundation established in Chapter 4, Chapter 5 examines how the cumulative pressure-induced stress (CPIS) affects models used to predict pipe breaks. Earlier research employed statistical approaches to evaluate CPIS's explanatory power for pipe breaks, identifying a positive correlation but yielding inconclusive results due to the observational nature of the data and constraints of the linear model, thereby prompting further study. In this chapter, CPIS data from a previously investigated, densely monitored network in the UK is analysed to explore the impact of CPIS on machine learning models for pipe break prediction. These models aim to address the limitations of earlier linear models and enhance the interpretation and practical use of CPIS. The results indicate that incorporating CPIS significantly strengthens model performance, suggesting that pressure fluctuations, and their characterisation by CPIS, are key factors underlying pipe breaks.

The findings of the previous chapter reinforce the assumption that cumulative pressure-induced stress (CPIS) can serve as a viable explanatory factor for pipe breaks and thus merits testing in large-scale, long-term experiments. However, to facilitate such experiments, the methods required for CPIS calculation must be revisited and refined from their

initial formulations to ensure both feasibility and reliability in experimental and practical applications. The remainder of this thesis investigates three technological prerequisites for accurately computing and applying CPIS at scale: (1) determining appropriate sampling frequencies and monitoring periods; (2) interpolating CPIS across networks with limited sensor availability; and (3) localising transient events to enable the correct implementation of the interpolation approach.

Chapter 6 addresses the first technological requirement by examining the sampling rate and monitoring duration necessary for cycle counting, based on a unique two-year dataset of high-resolution pressure measurements. The primary goal is to determine whether the sampling rate can be reduced while still capturing the cyclic loading induced by transient pressures in water networks. Reducing the sampling rate would prolong the lifespan of battery-powered pressure monitoring devices, thus lowering costs for continuous high-frequency monitoring. The findings indicate that a sampling rate up 10 S/s provides a satisfactory balance between measurement accuracy and energy efficiency.

After determining a suitable sampling rate and collecting high-resolution pressure data, the CPIS can be calculated at each monitoring point. The next technological challenge involves interpolating CPIS values throughout the network. Previous studies relied on networks with dense sensor coverage and linear interpolation methods, but questions about the suitability of these linear approaches have prompted further investigation into automated techniques for CPIS interpolation.

Chapter 7 presents a method for interpolating a "dynamic pressure" metric, which forms part of the CPIS, from an identified source of transients. Simulating transient pressures in water networks is both complex and computationally intensive. CPIS calculations rely on pressure data collected over periods ranging from weeks to years. Therefore, a simplified method is needed to capture the essential behaviour of transients over extended timescales. In this chapter, a technique is introduced that uses a brief, short-term transient simulation of an approximate hydraulic model, coupled with spatial interpolation methods, to accurately interpolate the dynamic pressure component of the CPIS. This

method is tested and validated using data acquired from a real-world operational case study.

The proposed CPIS interpolation method relies partly on identifying the source of pressure transients. However, pinpointing these sources is non-trivial, necessitating advanced methods for pressure transient source localisation to ensure that the automated interpolation approach remains viable. Therefore, Chapter 8 focuses on developing a methodology for locating pressure transient sources in water distribution networks. A key feature of this proposed method is its recognition of network connectivity uncertainty, caused by commonly unknown status of valves that can compromise transient localisation performance. Formulated as a mixed-integer program, the method integrates the k-shortest path routing problem to account for multiple possible wave propagation paths. It systematically evaluates potential source locations among all network nodes, considering every possible route. The performance of this approach is tested using both simulated and real data from an operational network, alongside an analysis of the minimum sensor density required for reliable source localisation.

Collectively, the preceding chapters establish a solid basis for implementing and testing the CPIS concept in large-scale, long-term experiments. These experiments are essential for demonstrating the reliability of CPIS as a metric to guide pressure management strategies aimed at extending the service life of ageing water distribution pipes. Chapter 9 consolidates the key accomplishments of this thesis, presents its main conclusions, and outlines future research directions.

1.5 Research dissemination

The work produced from this research has contributed to the following list of research publications:

1.5.1 Peer-reviewed journals

1. **Jara-Arriagada, C**, & Stoianov, I. (2021) Pipe breaks and estimating the impact of pressure control in water supply networks. *Reliability Engineering & System Safety*. 210, 107525, ISSN: 0951-8320

2. **Jara-Arriagada, C**, & Stoianov, I. (2024) Pressure-induced fatigue failures in cast iron water supply pipes. *Engineering Failure Analysis*. 155, 107731, ISSN: 1350-6307

3. **Jara-Arriagada, C**, Ulusoy, A.J., & Stoianov, I. (2024) Localization of Transient Pressure Sources in Water Supply Networks with Connectivity Uncertainty. *Journal of Water Resources Planning and Management*. 150(4), 107731, ISSN: 0733-9496

4. Konstantinou, C., **Jara-Arriagada, C**, & Stoianov, I. (2024) Investigating the Impact of Cumulative Pressure-Induced Stress on Machine Learning Models for Pipe Breaks. *Water Resources Management*. 38, ISSN: 603-619, 0920-4741

5. **Jara-Arriagada, C,** & Stoianov, I. (2025) Interpolating transient pressure metrics using sensor data for water distribution networks asset management. Under Review. *Water Resources Research*. ISSN: 0043-1397

6. **Jara-Arriagada, C,** & Stoianov, I. (2025) Field validation of a pressure transient source localization method in a water distribution network. Under Review. *Journal of Hydroinformatics*. ISSN: 1464-7141

1.5.2 Conference papers

1. **Jara-Arriagada, C**, & Stoianov, I. (2022) High resolution water pressure monitoring for the assessment of fatigue damage in water distribution pipes. *Proceedings of the 2nd International Joint Conference on Water Distribution System Analysis (WDSA)& Computing and Control in the Water Industry (CCWI)*. 1220-1235, ISBN: 978-84-9048-982-6

2. **Jara-Arriagada, C**, & Stoianov, I. (2023) Analytical evaluation of an optimization-based transient source localization algorithm under uncertainties in water supply networks *19th Computing and Control for the Water Industry Conference (CCWI) 2023*.

3. **Jara-Arriagada, C**, & Stoianov, I. (2024) Experimental evaluation of a pressure transient source localization method in an operational water distribution network. *15th International Conference on Hydroinformatics (HIC 2024)*.

Chapter 2

Literature review

This chapter reviews the most significant literature that precedes the research presented in this thesis. The main aim of this review is to contextualise and inform the research gaps that motivate the contributions presented in the following chapters. The literature review is organised from the foundational studies that set the background for the cumulative pressure-induced stress to the technological challenges for its implementation in real operational networks.

2.1 Pipe break analysis in the water industry

Proactive strategies for mitigating pipe breaks have been researched for decades, but widespread implementation remains limited. Research has primarily focused on developing predictive models for pipe breaks (Taiwo, Ben Seghier & Zayed, 2023), with the aim of prioritising resources for those pipes most susceptible to failure and carrying critical consequences within risk-based frameworks (Giustolisi et al., 2006). Water distribution pipes break when loading conditions exceed the material's tensile strength or the material's fracture toughness, initiating cracks that evolve into breaks. Corrosion processes can further exacerbate these failures by diminishing wall thickness or producing corrosion holes under specific conditions (Mazumder et al., 2019). In principle, a fully monitored

pipe enables the estimation of its deterioration and remaining service life through physical models, ranging from finite element analyses to simplified formulas, provided that all relevant loading and material properties are known (Rajani & Kleiner, 2001; Fu et al., 2020; Li et al., 2021). However, the extensive data required for these models is frequently costly and difficult to acquire, limiting such in-depth assessments only to critical transmission mains (Delnaz et al., 2023).

When analysing large cohorts of pipes where detailed examination is impractical, one of the most effective methods for estimating the likelihood of pipe breaks and remaining service life is through the use of statistical or machine learning models (Delnaz et al., 2023). These models require robust datasets encompassing both historical pipe breaks and potential covariates influencing failure events to yield statistically meaningful results (Taiwo, Shaban & Zayed, 2023). Because such studies rely on observational data, it is essential to apply engineering judgment and rigorous theoretical principles to reliably interpret the obtained results between the data and historical pipe breaks (Zhao & Hastie, 2021). This approach enables the identification of the primary factors governing the "epidemiology" of pipe breaks and helps pinpoint areas in the water distribution network that may demand targeted interventions to mitigate future failures.

Multiple statistical and machine learning methods have been employed to model pipe breaks (Dawood et al., 2020), with the choice of technique largely depending on the nature of the available data (Nishiyama & Filion, 2013). Scheidegger et al. (2015) distinguish between models that utilise data specific to individual pipes and those based on broader network-level characteristics. In the water industry, models incorporating pipe-level data are typically preferred, owing to the substantial heterogeneity of pipes and loading conditions that can coexist within a single network.

A common characteristic among statistical models for pipe break analysis is that, regardless of their formulation, they seek to fit a predefined failure rate function (Scheidegger et al., 2015). The advantage of this strategy is that statistical tests can be used to verify whether the assumptions underpinning the failure rate function are valid. However, this

inherent rigidity may limit the detection of potential non-linear effects not captured by the model's formulation. In contrast, machine learning models offer greater flexibility in data fitting, albeit at the risk of overfitting or identifying spurious relationships (Andaur Navarro et al., 2021).

Despite the potential drawbacks of models for pipe break analysis, these models and available data have led to a relatively good understanding of the primary factors influencing pipe breaks (Barton et al., 2019). For instance, several studies have demonstrated the significant impact of corrosion on pipe deterioration (Hussein Farh et al., 2023). Corrosion patches reduce the wall thickness of pipes, compromising their structural integrity and resistance (Fu et al., 2020). Additionally, localised corrosion can create holes and cracks that significantly weaken pipes to normal operational loads, further reducing their durability (Liyanage & Dhar, 2017).

Temperature is another factor strongly linked to pipe breaks, particularly during winter (Rajani, Kleiner & Sink, 2012). Studies have shown a strong correlation between pipe breaks and low temperatures potentially due to the freeze-thaw cycles that can exceed the tensile strength of pipes, especially in areas weakened by corrosion (Habibian, 1994; Wols & Van Thienen, 2014). Moreover, hot weather conditions, which can cause ground movements, are increasingly associated with the failure of asbestos cement pipes, which are sensitive to such stresses (Wols & Van Thienen, 2014). Both extreme hot and cold weather conditions, exacerbated by climate change, are expected to continue imposing greater loads on pipes, potentially leading to some inevitable pipe breaks.

Geometric factors intrinsic to individual pipes have shown to also play a significant role for the estimation of breaks. Notably, smaller diameter pipes are more prone to breakage than larger ones (Pelletier et al., 2003; Laven et al., 2024). This seems to be consequence of their thinner walls, which make them more susceptible to corrosion, bending, and also potential overpressures caused by transient reflections, as these pipes usually serve as dead ends of branched networks (Pelletier et al., 2003).

A common characteristic of the aforementioned factors influencing pipe breaks is that they

lie outside the direct control of water utilities. Moreover, the extensive and heterogeneous nature of pipes in water networks frequently provides minimal insight into prioritising areas of concern, even when statistical models are employed.

Another critical factor influencing pipe breaks is the water pressure within the distribution system. Unlike many other variables, water pressure can be partly controlled by water utilities (Wright et al., 2015; Martin et al., 2022), making it particularly relevant. This consideration suggests an additional application for predictive pipe-break models: assessing how water pressure management might extend the life of ageing infrastructure. However, despite the potential importance of water pressure, its precise role in pipe breaks remains poorly understood (Barton et al., 2019). This knowledge gap largely stems from limited pressure monitoring across networks and the lack of reliable, calibrated hydraulic models that can accurately determine pressure levels at specific locations (Barton et al., 2020). As a result, most studies have found only weak evidence of water pressure's influence or have merely identified correlations without exploring how pressure management could reduce pipe breaks (Winkler et al., 2018).

For instance, Fan et al. (2022) modelled pipe breaks using pipe elevation as a substitute for water pressure due to a lack of pressure data. Their findings showed that pipes in higher elevations were less prone to failure, likely due to lower service pressure. Similarly, Rifaai et al. (2022) used pressure zones in their pipe break analysis as a proxy for water pressure but reported an unexpected negative coefficient, without offering a discussion for this contradictory result. Winkler et al. (2018) were also constrained by the lack of pressure data to model pipe breaks. They had to use the pipes' nominal pressure to assess the effect of water pressure, which their results showed to be the least significant factor among the used covariates (Winkler et al., 2018). They acknowledged that using nominal pressure was insufficient to capture the true impact of water pressure on pipe breaks, as reflected in their findings (Winkler et al., 2018). In these studies, water pressure was only one of many factors used to fit pipe break models. Other studies have focused exclusively on water pressure.

2.2 Water pressure and pipe breaks

Some studies have provided exclusive attention to the links of water pressure and pipe breaks with modern pipe break datasets and pressure metrics, motivating further research. for instance, Martínez-Codina et al. (2016) implemented an statistical analysis to assess differences in the cumulative distribution function (CDF) of pressure metrics of pipes conditioned to break against pipes unconditioned to break. It was observed that CDFs of pressure range, which is the difference between maximum and minimum pressure, were significantly different when estimated for pipes conditioned to break against pipe unconditioned to break (Martínez-Codina et al., 2016). This result suggests that pressure range has a potential influence on the pipes that suffered breaks. Maximum average pressure was also assessed, however, it was not a significant variable for the occurrence of pipe breaks. A possible explanation was that the maximum pressure was not high enough so that it was not significantly different between pipes that broke and those that did not break.

In another study with a wider range of pressure data, Martínez García et al. (2020) performed a spatiotemporally based clustering analysis to assess the correlation of consistent high pressure with pipe failure rates. Results from this study showed that there was a strong correlation between high pressure and pipe breaks in areas with high failure rates. However, weak correlations were also observed in other areas, suggesting that the effects of pressure could be spatially dependent.

In addition to academic studies, early industry experience reported by Pearson et al. (2005), and Fantozzi et al. (2009, 2013) presented that pressure management have been correlated with significant reductions in pipe breaks. For example, Pearson et al. (2005) presents a case study where, following the implementation of pressure management, the monthly frequency of pipe breaks dropped by over 70% in a combined dataset of reported pipe breaks from multiple water utilities across various countries. Likewise, Fantozzi et al. (2009) and Fantozzi et al. (2013) also reported immediate reductions in pipe breaks following pressure management.

The mechanical explanation for these results was suggested on the basis of an empirical pipe failure criterion provided by Schlick and revisited by Rajani et al. (2000). The model considers the effect of stresses produced by internal water pressure and traffic loadings to create a parabolic failure envelope for the pipes (Rajani et al., 2000). Then, it is easy to show that by reducing operational pressure, the failure point of a pipe moves to safer levels, extending the pipe's life until corrosion or other deterioration brings it back into a danger zone. Therefore, lowering average pressure can lead a pipe's performance to a younger, less deteriorated state.

However, despite these reported successes, the mechanical rationale and recent academic works, many pipe breaks continue to occur, and pressure management has not been widely accepted in the water industry as a solution to reduce pipe breaks. This reluctance may stem from a variety of factors, including the need to explain instances where pressure management did not lead to reduction in burst frequency. A suggested candidate to explain those occurrences was the impact of pressure fluctuations, and the additional failure mechanisms involved in pressure management that may influence the relationship between pressure and pipe breaks.

2.3 Pressure fluctuations and fatigue failure of pipes

Pressure fluctuations occur when the speed of water flowing through pipes changes quickly, creating pressure waves that move back and forth until they dissipate. These fluctuations, also called pressure transients, reflect the transition from one hydraulic state to another (changes in pressure) caused by the sudden velocity changes (Karney & McInnis, 1990). Typically, these sudden changes in flow velocity are caused by operations in water networks, such as adjusting valves or running pumps (Karney & McInnis, 1990). Additionally, changes in water demand and pipe breaks can also generate these waves, which can produce additional loading conditions to pipes, including impact loading and fatigue (McInnis & Karney, 1995; Rezaei, 2017).

It was previously believed that pressure fluctuations in water networks were rare, largely due to limited monitoring technologies capable of measuring pressure and flow at high enough resolutions to detect pressure transients (Rezaei et al., 2015; Rathnayaka et al., 2015). The speed of pressure transient propagation depends on factors such as pipe material, flow conditions, and the anchoring system, with pipe material being the most significant in water networks (Ellis, 2008). Pressure waves can travel from around 300 m/s in plastic pipes to as fast as 1300 m/s in cast iron pipes (Covas et al., 2004; Twyman, 2016). These high propagation speeds exceed the ability of typical supervisory control and data acquisition (SCADA) systems to capture pressure fluctuations accurately (Tornyeviadzi & Seidu, 2023). Additionally, it was assumed that pressure transients would dissipate quickly if they occurred, drawing little concerns about their impact on the system (Starczewska et al., 2014).

However, advances in high-speed pressure monitoring began to reveal that pressure fluctuations are much more frequent than previously thought (Stoianov et al., 2007; Hoskins & Stoianov, 2014). Monitoring campaigns have shown that industrial and commercial users generate large pressure transients in their routine operations, making them a continuous source of pressure fluctuations (Hoskins & Stoianov, 2014; Starczewska et al., 2015). Additionally, efforts to reduce leakage through pressure management strategies, such as the use of pressure control valves, have increased operational activities, further contributing to the occurrence of pressure transients (UKWIR, 2022). This discovery of ongoing dynamic pressure activity in water networks through advanced monitoring has raised concerns about the potential for cyclic loading from these fluctuations to cause fatigue in pipes (UKWIR, 2022). This could reduce pipe strength over time, eventually leading to failure under normal operating conditions. These findings introduce a new approach to explain the correlations observed in areas with high pressure variations and pipe breaks, which could in reality be areas with high pressure fluctuations affected by fatigue if measured at high resolutions.

Early evidence of high-resolution water pressure snapshots, such as those presented by McInnis (2004) and Stoianov et al. (2003, 2007), made clear that pressure transients could

impact pipe integrity. However, the lack of technological developments to perform large scale studies to confirm the widespread of pressure transient limited further research. Motivated by the need of better technological developments, Hoskins (2015) initiated research to develop a robust data acquisition system and analytics to assess the impact of pressure fluctuations on pipe integrity. The results from Hoskins (2015) led to the development of advanced wireless high-speed pressure monitoring devices, as well as methods for managing and analysing large datasets generated by these devices (Hoskins & Stoianov, 2014). The technology was applied in a large observational program, which confirmed the widespread presence of pressure transients in water networks and highlighted the need for further investigation into their effect on pipe integrity.

Similarly to the work of Hoskins (2015), Starczewska (2016) also studied the impact of pressure transients on pipe integrity by gathering data in another large-scale experiment that deployed commercial high-speed pressure sensors for 11 months, with a total of 144 locations monitored. The results of such observational study further confirmed that several pressure transients with range of more than 30 mH2O can occur every day in various regions of water networks. A methodology called transient fingerprint was develop to characterise the recorded transient pressure data. However, the method did not allow to produce an extrapolation of the measured locations limiting their ability to assess the impact of pressure transient with pipe break records to assess pipe integrity.

Following the line of research of Hoskins (2015), Rezaei (2017) developed another large scale observational study to characterise pressure dynamics with pipe breaks. A total of 480 locations were monitored with a minimum monitoring period of 2 weeks. The main goal of the research of Rezaei (2017) was to expand the evidence of hydraulic pressure instabilities in water networks and assess their influence on pipe breaks. To this end, the basics of fatigue failure of materials was reviewed with the aim to provide a new metric to assess the impact of pressure dynamics on pipe breaks. This initial investigation led to the enhancement of a novel concept proposed in a patent by Hoskins & Stoianov (2017), termed cumulative-pressure induced stress. This concept aims to explain the failure of water pipes by the combined effect of long-term pressure fluctuations and the

mean pressure in pipes as presented in the next section.

2.4 The cumulative pressure induced stress

The cumulative pressure-induced stress (CPIS) has been proposed as a novel concept to assess the impact of pressure transients on pipe breaks Hoskins & Stoianov (2017). Motivated by common concepts of fatigue failure of materials, this metric was developed to consider the impact of cyclic loadings caused by pressure fluctuations and the mean pressure. The main drive for developing the CPIS metric arose from the insufficient attention given to monitoring pressure transients, and particularly transients with medium to small amplitude (Rezaei, 2017). Pressure transients of large magnitude can potentially break a deteriorated pipe by exceeding the tensile strength of pipes. Nevertheless, it is hypothesised that a main drive that form these deteriorated pipes, alongside corrosion is the accumulation of frequent pressure fluctuations. According to basics of fatigue of materials, several materials can be characterised by an S-N curve, which shows how the resistance of a material degrades from an original state to a more fragile state caused by cyclic loadings.

Motivated on these concepts, the original patent that presented the CPIS (Hoskins & Stoianov, 2017) showed the guidelines of the concept and development of the metric. The main concept of the CPIS was thought as a function of three main variables as follows:

$$CPIS = f(\text{mean pressure}, \text{amplitude or range of pressure cycles}, \text{cycle counts}) \quad (2.1)$$

but the actual functional form for CPIS was left open for further research. The functional form of the CPIS was further detailed by Rezaei (2017), who presented CPIS as a metric composed of two variables as follows:

$$CPIS = f(P_{mean}, DP) \tag{2.2}$$

where, P_{mean} is the diurnal mean pressure in the pipe and DP, dynamic pressure, is a proposed metric that combines the range of pressure cycles and cycle counts as follows:

$$DP = \sum_{i=1}^{k} (S_i \cdot n_i) \tag{2.3}$$

where, k is the number of different pressure cycle events experienced by a pipe, S_i corresponds to the range of the cyclic loading event i, and n_i is the number of pressure cycles with range i.

The specific functional form of the two variables comprising CPIS (Equation 2.2) was proposed to be implicitly learned when the variables are decoupled and included individually within a statistical or machine learning model (Rezaei, 2017). The development of the DP metric described by Rezaei (2017) presents, however, strong assumptions based on linear damage accumulation following Miner's rule (Miner, 1945). Furthermore, Rezaei (2017) assumes a stress to number of cycles relationship that is not observed in experimental data.

Miner's rule (Miner, 1945) was developed to evaluate the cumulative effects of cycles of varying stress levels applied to a material. It gained popularity because of its simple assumptions, but it has also faced criticism for this same reason. The fundamental idea of Miner's rule is that stresses at different magnitudes accumulate linearly over time, consuming part of a material's fatigue life. If there are k distinct cycle stress levels, S_i, applied to a material, and the average number of cycles to failure at the $i-th$ stress level is N_i then the damage fraction, D_f, experienced by the material is given by:

$$D_{fraction} = \sum_{i=1}^{k} \frac{n_i}{N_i} \tag{2.4}$$

When the $D_{fraction}$ reaches the value of 1, the material is expected to fail. The value of N_i for a particular stress level is obtained from a S-N curve. This curve is obtained experimentally, and indicates for different stress levels, the number of cycles needed for a component to break.

Another form to express linear damage accumulation is through a damage index (Rezaei, 2017; ReliaSoft Corporation, 2024), which can be calculated for a given stress level and the number of cycles applied under that stress as:

$$W = S_i \cdot n_i \tag{2.5}$$

If we assume a critical damage index at failure being the same for all stress levels (which is not observed in real fatigue experiments), we can express it as:

$$W_{failure} = S_i \cdot N_i \tag{2.6}$$

Then, for instance, if $W_{failure} = 100$ for an element, the element would fail after 20 cycles at a stress of 5, or after 50 cycles at a stress of 2 (ReliaSoft Corporation, 2024). If we multiply the original Miner's rule expression by S_i and introduce the concept of $W_{failure}$ and the DP metric:

$$D_{fraction} = \sum_{i=1}^{k} \frac{n_i \cdot S_i}{N_i \cdot S_i} = \frac{n_1 \cdot S_1}{N_1 \cdot S_1} + \frac{n_2 \cdot S_2}{N_2 \cdot S_2} + \cdots + \frac{n_k \cdot S_k}{N_k \cdot S_k} \tag{2.7}$$

$$\Rightarrow \frac{\sum_{i=1}^{k}(S_i \cdot n_i)}{W_{failure}} = \frac{DP}{W_{failure}} = D_{fraction} \tag{2.8}$$

We can observe that the dynamic pressure metric, DP, represents the portion of the critical cumulative damage index at failure under the assumption of linear damage accumulation (Rezaei, 2017). However, as presented in this section, the assumptions underlying

DP present notable limitations. As mentioned by Rezaei (2017) the CPIS (DP) metric does not account for the effect of loading sequence, an inherent limitation of Miner's rule. Although Miner's rule has been criticised for its assumption of linear damage accumulation, it is considered a conservative approach that provides a safety margin in practical engineering applications.

The second major assumption in the DP formulation is more problematic. The assumption of an equal damage index at failure across all stress levels is not consistent with most S-N curves. S-N curves are typically best modeled using exponential relationships, such as $C = S \cdot N^m$, with the Basquin equation as a common example (Burhan & Kim, 2018). The multiplicative form of the DP metric, which lacks an exponent to reflect differences between low-cycle and high-cycle fatigue, may hinder an accurate association of CPIS with fatigue failure in pipes. Specifically, the DP calculated at low stress levels to break a pipe would result in a much larger number than the DP calculated from higher stress levels. In practical applications, it might result in effects of higher stress cycles being masked by the larger number of lower stress cycles, reducing the power of the DP variable.

Despite these limitations, Rezaei (2017) assessed the impact of CPIS on pipe breaks, using DP and mean pressure, by developing four datasets based on extensive data from a large observational experiment. Their analysis, using linear logistic regression, found a positive association between DP and pipe breaks. An increase in DP was consistently linked to a rise in pipe breaks across all models. The results for the best-fitting model indicated that the probability of pipe breaks increased from 0.12 at the minimum DP level to just under 0.2 at the maximum DP level in the dataset.

However, another potential limitation of Rezaei (2017)'s approach is the linear nature of the logistic regression model they employed. The model did not include quadratic or interaction terms, which could have captured non-linear relationships among the variables. Additionally, while DP was found to be positively associated with pipe breaks, there was limited theoretical justification for the use of fatigue metrics to explain failures in water pipes. Lastly, many of the methods used to estimate CPIS were still in early stages of

development, reducing their practical applicability, especially in terms of interpolating cycle count metrics, such as the DP, for all pipes across a network. Despite the several limitations, CPIS, and the use of the DP metric found positive associations with pipe breaks, motivating further investigations into the suitability of this concept for pipe break analysis.

2.5 High frequency pressure monitoring

To characterise and evaluate the impact of the cyclic loadings caused by pressure transients, high resolution water pressure monitoring is required. Traditionally, water pressure has been measured using Supervisory Control and Data Acquisition (SCADA) systems at sampling intervals ranging from 15 to 5 min. While this resolution is sufficient for characterising quasi-steady state water pressure in hydraulic control and modelling, it falls short for capturing pressure dynamics.

Pressure transients are essentially pressure waves that travel through pipes at high speed. To accurately reconstruct these waves, high-frequency pressure monitoring is needed. According to Hoskins (2015), the fundamental frequency of these waves is around 5-10 Hz (excluding those caused by cavitation). On the basis of the Nyquist rate, a principle in signal processing, the minimum sampling rate required to capture the frequency components of a signal is twice its fundamental frequency, about 10-20 samples per second in this case. However, to analyse the impact of pressure waves, a much higher sampling rate is necessary to capture both the wave's shape and its maximum amplitude. In practical terms, this requires sampling frequencies in the range of 50-100 samples per second Hoskins (2015). With this rate, and assuming an average wave speed of 1000 m/s, the transient pressure shape can be reconstructed with a resolution of 20 to 10 m Hoskins (2015).

Once pressure transient waves are characterised in a pressure time series, their impact is typically assessed by counting the number of cycles at different pressure ranges. This

process, called cycle counting, can be performed using various algorithms (Cheng & Broz, 1986; ASTM, 2017). As previously mentioned, higher sampling rates improve the accuracy of pressure wave reproduction and therefore results in cycle counting are also improved. However, high sampling resolution significantly increases the energy consumption of battery-operated pressure transducers (Hoskins, 2015). This highlights the need to determine an optimal sampling resolution that is sufficient to capture the maximum pressure amplitude while minimising energy use.

In addition to the complexities of power consumption, high-resolution pressure monitoring also presents challenges in data management and processing. The volume of data generated from just a few hours of high-resolution monitoring can exceed gigabytes, requiring significant storage, analysis, and processing capacity. According to the findings of (Rezaei, 2017), a two-week logging period is recommended to provide a reliable representation of a full year's pressure cycle activity in networks with minimal annual variations in user demand and hydraulic control operations. This period, however, still presents a challenge for the practical application of CPIS.

2.6 Technological challenges for CPIS interpolation

The cumulative pressure-induced stress (CPIS) concept has been proposed to consist of a combination of the mean daily pressure in pipes and a metric that represents the effect of cyclic loading. Metrics developed to characterise this concept can be correlated with historical pipe breaks to assess the impact of pressure fluctuations. For such correlations, the metrics generated through the CPIS concept must be linked to each individual pipe in the network. Mean daily pressure is relatively easy to estimate using conventional fifteen-minute pressure readings from SCADA systems and extended period hydraulic simulations, which are calibrated with those readings. However the estimation of metrics related to pressure transients is far more complex.

A metric to characterise the impact of pressure dynamics requires counting the number of

cycles produced by pressure transients. This process involves estimating the propagation of pressure transients and their reflections over a time period long enough to fully capture their effects. However, current capabilities for simulating pressure transients are limited. Several methods exist for analysing transient propagation, with the most common being the method of characteristics and the wave characteristic method (Ramalingam et al., 2009). While these methods accurately represent the rising edge of pressure waves generated by transient events, they often struggle to accurately model the reflections of these waves (Vítkovský et al., 2000; Plouraboué, 2024). These methods also need to establish the sources of pressure transients and boundary conditions. Furthermore, they remain computationally intensive, limiting their use to small networks and short simulation periods, which restricts their scalability for CPIS computations. Based on the experience of Rezaei (2017), collecting high-resolution pressure data for a period of about two weeks would be necessary to adequately capture the pressure dynamics in water networks. However, simulating even one day of transient pressure data with current analytical methods is impractical given present computational capabilities.

Alternative approaches to transient pressure simulations for estimating pressure metrics rely on regression or interpolation methods. In his analysis of the DP metric, Rezaei (2017) proposed a regression model to estimate DP values between nodes in a network. This model uses covariates calculated along the path between source and target nodes such as the percentage of metallic pipes, the percentage of pipes with diameters smaller than 250 mm, demand flow, and the number of customers. However, this approach is highly simplified and overlooks several key factors involved in pressure transient propagation and reflections. Such factors involve dead ends, reflection points, friction factors, and pipe lengths. Despite these limitations, the model was applied to a densely monitored network, which helped reduce uncertainty in the results. In real-world applications, however, networks may not be monitored as intensively as in Rezaei (2017)'s study, and the method lacks validation, further limiting its practical use.

Another approach to estimating pressure transient metrics is the use of interpolation methods, commonly applied in the analysis of geo-spatial data (Li & Heap, 2014). The

use of these methods would assume that the transient behaviour and the number of cyclic pressures measured by sensors are influenced by the network's location being analysed. While this assumption is reasonable, it must be tested, especially if pressure transient sources are assumed to remain stable over long periods. Commonly used interpolation methods for spatial data include inverse distance weighting (IDW) (Shepard, 1968), and Kriging (Chilès & Desassis, 2018), which relies on spatial autocorrelation. The choice of method depends on the type, quantity, and quality of the available data. Therefore, evaluating the suitability of different interpolation methods for pressure data is essential.

Given the complex challenge of interpolating pressure metrics in water networks, a hybrid approach that combines transient simulation with interpolation methods is envisioned. This approach would involve identifying the sources of pressure transients, which is a required step for accurate transient simulations. Additionally, locating sources of pressure transients makes possible to mitigate the impact of hydraulic instabilities in water networks using transient control devices (Boulos et al., 2005).

2.7 Localisation of sources of pressure transients

A prevalent method found in existing literature for localising the sources of pressure transients relies on analysing the differences in the arrival times of the first wave to pairs of time-synchronised sensors, before any reflections occur (Silva et al., 1996; Stoianov et al., 2007; Srirangarajan & Pesch, 2013; Guibert et al., 2023). When a transient event occurs, the resulting pressure wave propagates through the pipes and arrives at time-synchronised pressure monitoring devices at different times, which depend on the distance from the source of the transient and wave velocity (pipe properties). By analysing the time delays between the arrival of the first pressure wave at each monitoring device, it is possible to triangulate the location of the most probable source of the pressure transient event. Silva et al. (1996) investigated this approach on an experimental setup consisting of a single pipe connected to four pressure transducers, and a valve to generate

fast pressure transients. The authors successfully detected the source of a transient by analysing the time differences of the first pressure wave arrival times to the different pairs of time synchronised pressure transducers. Misiunas, Vítkovský, Olsson, Simpson, Asce & Lambert (2005) applied this idea and investigated the use of a single pressure monitoring device to detect the source of a transient by considering also the time of wave reflection in an individual pipeline under controlled laboratory conditions. Even though the method looks promising for a single pipeline, the complexity of analysing wave reflections in complex networks can undermine its efficacy, rendering this approach unsuitable for operational networks.

The approaches discussed above primarily focused on analysing sources of transients in individual pipelines. Srirangarajan et al. (2010) expanded this concept to a network by using a scoring algorithm. As in previous studies, the arrival times of the first pressure wave at different time-synchronised monitoring locations in a WSN are recorded. Then, given knowledge of the pressure wave speeds derived from the pipe materials and pipe lengths, the wave fastest paths from every node in the network to the sensor locations are calculated using graph theory. With this information, the analytical time differences for the first pressure wave reaching different sensors from every potential source node are calculated. These differences are then compared to the actual measured differences in arrival times (Srirangarajan et al., 2010; Srirangarajan & Pesch, 2013). The source of the transient is determined by the node that scores the closest alignment between differences in the measured and analytical arrival times of the pressure wave to the sensors (Srirangarajan et al., 2010; Srirangarajan & Pesch, 2013). For this procedure to work, at least two sensors are required to determine the time differences of arrival times between sensors. This concept was experimentally validated in a laboratory setup by Hampson (2014) and also applied by Kang et al. (2018) to localise bursts that cause sudden pressure drops. Recently, Zecchin et al. (2022) investigated the optimal sensor placement in order to optimise the detection and localisation capabilities of the timing-based pressure transient source localisation method.

Previously explored algorithms for localising the source of a pressure transient, which use

timing-based scoring, generally operate under the presumption that the water network's connectivity is known and accurate. However, hydraulic models of operational networks include uncertainties in the status of valves and network connectivity. When applying a timing-based algorithm to assess the paths of pressure wave propagation in networks with uncertain connectivity, mismatches between the actual network and the hydraulic model can occur. These discrepancies can lead to the failure of the algorithm.

2.8 Summary

This chapter has briefly reviewed relevant previous research with the aim of identifying research gaps in the utilisation of the CPIS concept for pipe break analysis. This review has drawn several conclusions which motivate the research objectives, contents and contributions of the following chapters of this thesis as presented below:

- Previous studies that have analysed the effect of water pressure on pipe breaks have primarily focused on identifying significant coefficients in statistical models. However, they have not explored how reducing pressure metrics, such as mean pressure and pressure range, impacts pipe breaks. Additionally, data in many studies has been limited, and the models used (linear or machine learning) have not been thoroughly evaluated for physical coherence. In this context, and before further investigating the impact of pressure fluctuations on pipe breaks, Chapter 3 addresses the previous issues by conducting a more comprehensive statistical analysis with a larger dataset and ensuring physical coherence among covariates and pressure variables to strengthen the reliability of the findings. Results of this more detailed approach would reinforce findings about impact of water pressure components on pipe breaks. In particular, if pressure variability is strongly related to pipe breaks as in previous studies, it would reinforce the analysis of pressure fluctuations, and the association of fatigue failure.

- Hoskins & Stoianov (2017) and Rezaei (2017) introduced the concept of cumulative

pressure-induced stress (CPIS) to assess the impact of pressure dynamics on pipe breaks. This concept is based on the hypothesis that pressure fluctuations, as recorded by high-speed pressure loggers, are frequent and large enough to cause fatigue failure in pipes. Despite this hypothesis seems plausible, it has not been rigorously verified through technical or theoretical analysis, leaving uncertainty about the actual importance of pressure fluctuations. To address this gap, Chapter 4 examines the fundamentals of pipe fatigue failure from a fracture mechanics perspective, which is more appropriate for deteriorated pipes than conventional fatigue analysis. The main objective is to apply fracture mechanics principles to determine the conditions under which pressure fluctuations can cause fatigue failures on aged deteriorated pipes.

- Rezaei (2017) investigated the impact of CPIS on pipe breaks using linear logistic regression, finding a positive correlation. However, the lack of theoretical validation, the restrictions of the linear model and weaknesses in the CPIS formulation demand more research to validate previous findings. Chapter 5 revisits the CPIS metric and the dynamic pressure component (DP), aiming to improve the CPIS modelling approach through more flexible machine learning models. Additionally, a more robust framework for interpreting the results is developed.

- Calculating CPIS requires high-resolution pressure monitoring, but the extensive data collected through this method poses challenges in terms of energy consumption and storage. Chapter 6 addresses this issue by evaluating the optimal sampling resolution and minimum monitoring period needed to reliably estimate CPIS components while minimising resource demands.

- Rezaei (2017) developed foundational methods for the extrapolation and interpolation of CPIS. However, this review identified that these methods still require further refinement to be effectively applied in practical real-world scenarios. This review has highlighted that pressure transient analysis still presents a computational challenge for analysis and simplified methods are still needed for practical purposes.

Furthermore, it was presented that geospatial interpolation techniques hold significant potential for estimating CPIS components across entire pipe networks given their simplicity. To better improve extrapolation/interpolation methods for CPIS, in Chapter 7 a novel hybrid approach is proposed, which integrates simplified transient pressure modelling with interpolation techniques to enhance accuracy in predicting CPIS values throughout the network.

- Methods for CPIS estimation require analysing and modelling pressure transients, which depend on knowing boundary conditions and, more importantly, pressure transient sources. Various methods in the literature use the arrival times of pressure waves detected through high-resolution monitoring to pinpoint the location of transient pressure sources. However, these methods would struggle with uncertainties in network connectivity, pipe wave speeds, and determining wave arrival times accurately. To improve the reliability of pressure transient source estimation, Chapter 8 reformulates earlier pressure transient source localisation methods as a mixed-integer optimisation problem. This approach considers multiple wave propagation paths to account for connectivity uncertainty and includes constraints to ensure the solution is close to the actual transient source. In this way, CPIS estimation can be more reliably calculated, and the located sources of transients can be managed to reduce the impact of pressure transients.

Chapter 3

Average water pressure metrics and pipe breaks

Related publication: Jara-Arriagada, C, & Stoianov, I. (2021) Pipe breaks and estimating the impact of pressure control in water supply networks. *Reliability Engineering & System Safety.* 210, 107525. doi:10.1016/j.ress.2021.107525

3.1 Introduction

The literature review has outlined the background of cumulative pressure-induced stress as a factor contributing to pipe breaks in water distribution networks. This concept originated from recurrent pressure transients observed in monitoring campaigns of pressure measured at high resolution. Furthermore, early evidence of past research has shown that steady state water pressure is related to pipe breaks, but not consistently observed in industry practice. This phenomenon is potentially explained by the interaction of steady state pressure levels and pressure fluctuations causing fatigue. However, technical limitations in measuring pressure with sufficient temporal and spatial resolution have hindered the ability to accurately evaluate the potential of such hypothesis. The literature review showed that the CPIS requires refined investigation in three key areas of motivation,

theoretical foundations, and methods for evaluation.

As a first step to initiate a refined investigation of the CPIS concept, this chapter regress to investigate and confirm the evidence of steady state water pressure metrics influencing pipe breaks. Modelling approaches using historic pipe breaks have shown the statistical significance of water pressure on pipe breaks. However, these models have not been evaluated for physical coherence and the interaction of water pressure and other covariates. The primary objective of this chapter is to verify relationships observed in previous studies and assess the pressure levels associated with frequent pipe breaks in order to draw conclusions about the physical processes underlying these breaks, and could potentially justify the analysis of pressure dynamics (fluctuations).

To achieve these objectives this chapter uses an extensive dataset of pipe breaks, pressure data and pipe break explanatory variables collected from a water utility in the UK. The pressure data was obtained from the steady state hydraulic model used by the water utility for design and strategic planning. The pipe break records span a period of approximately 25 years, and a set of additional explanatory variables for pipe breaks are combined to produce an statistical model using machine learning techniques to achieve a non-linear fitting. In this way limitations of previous linear models are overcome. The model used for fitting pipe breaks records was based on a logistic regression with polynomial terms. The inclusion of these terms allows the model to explore interactions and non-linear relationships, which could not be fitted with a standard logistic regression. The large extend of the dataset, allowed to develop large sub datasets of pipe breaks of cast iron pipes and asbestos cement pipes. Furthermore, these datasets were further sub-sampled into seasonal pipe breaks of winter and summer for a more detail evaluation of the impacts of water pressure.

The large dataset employed allowed a detail exploratory analysis of the main characteristics and trends of the pipe breaks observed by the water utility. The results of the modelling approach were used to produce sensitivity analysis plots to evaluate to what extent predicted pipe breaks may vary with changes in average metrics of water pres-

sure. In this study the focus was on assessing average mean pressure in the pipes and the pressure range (difference of maximum vs minimum pressure). All results show that achievable reduction of mean pressure of about 20 m can significantly reduce about 30% of pipe breaks in this case study. Furthermore, the most impactful result is the association of pressure range and pipe breaks being more significant. However, despite this result, it was observed that the minimal changes in water pressure range seem excessive to produce such large modelling reductions on pipe breaks, suggesting that the pressure range variable could be a potential surrogate variable for dynamic pressure variations (fluctuations), which form part of the foundations of the cumulative pressure-induced stress concept.

3.2 Methods

This chapter investigates the relationships between average metrics of water pressure and pipe breaks. The aim of this investigation is to quantify the relevance of mean pressure and pressure range on pipe breaks as presented in past research. Furthermore, on the basis of the obtained results, this chapter aims to draw hypotheses with respect of the influence of pressure fluctuations on pipe breaks and therefore CPIS.

The analysis relies on the use of a logistic regression model with polynomial terms (or features) to predict pipe breaks. After an assessment and validation of the algorithm for the considered case study, a sensitivity analysis was implemented to assess variations in predictions against changes in the investigated water pressure metrics.

3.2.1 Data preparation

Historic records of pipe repairs for a period of approximately 25 years were received from a UK water utility. The pipe repairs records are an approximation for pipe breaks; therefore, there is inherent uncertainty with the time of occurrence of pipe breaks. The dataset contained missing values or inconsistent information, which is common for operational

data. These values were modified with *nan* or *unknown labels*. The exploratory analysis utilised all available records. However, for the predictive analysis, only entries of unique pipes were used that contained all necessary information. Also, as the dataset contained different pipe materials, these were not mixed for modelling, since asbestos cement and cast iron, for example, have different structural properties (Barton et al., 2019).

Before utilising the data in the models, log transformations were carried out for the variable length and the variable pressure range since these were right-skewed. Z-score normalisation was also implemented to normalise the range of the data for optimisation purposes.

3.2.2 Logistic regression

Logistic regression is a generalised linear model, which is commonly used for modelling data with binary dependant variables. It is a supervised learning algorithm for solving classification problems. Logistic regression was chosen for this study over alternative machine learning techniques for the following reasons: (1) It is easier to interpret even with the use of polynomials terms compared to other models such as neural networks. (2) The algorithm does not need to perform complex hyperparameter optimisations, which provides a more user-friendly and reproducible methodology for operators. (3) The use of polynomial terms expand the algorithm to capture non-linearities with a similar power to advanced machine learning methods. This is achieved, however, at the cost of being computationally expensive for large feature spaces.

The implementation of logistic regression for this study is as follows (Ng, 2007*b*):

Firstly, a dataset with m pipe records was compiled from the water distribution network. Each pipe record i in the dataset contained a pair $(x^{(i)}, y^{(i)})$, where:

$x^{(i)}$: was a vector of attributes, features or variables for the pipe i. This included, inter alia, the length of the pipe, diameter and year of installation.

$y^{(i)}$: was the state of the pipe i broken or no broken, as a binary outcome 1 or 0.

It was desired an algorithm capable of using the variables x of a pipe to provide the expected outcome y of the pipe as broken or not. Therefore, it was sought a function that outputs results in the range 0 to 1 given variables x that might be affected by parameters θ. To accomplish this, the sigmoid function is commonly utilised as a hypothesis function h_θ:

$$h_\theta = g(\theta^T x) = \frac{1}{1 + e^{-\theta^T x}} \tag{3.1}$$

where $g(z) = \frac{1}{1+e^{-z}}$ is called sigmoid function. In addition, formulating:

$$p(y = 1|x; \theta) = h_\theta(x) \tag{3.2}$$

$$p(y = 0|x; \theta) = 1 - h_\theta(x) \tag{3.3}$$

Equation 3.2 indicates that $h_\theta(x)$ is sought to represent the probability of the output y being equal to 1 given a vector of variables x parametrised by a vector of parameters θ. Equation 3.3 specifies that the probability of $y = 0$ should be equal to 1 minus the result of $h_\theta(x)$. Both expressions can then be condensed in one single expression called the Bernoulli distribution:

$$p(y|x; \theta) = h_\theta(x)^y [1 - h_\theta(x)]^{1-y} \tag{3.4}$$

Secondly, for our data set of m pipes, it can be assumed that every observation was generated independently and identically distributed. This allows the use of the method of maximum likelihood estimation to determine the value of the parameters that maximise the probability of the data. For example, find the parameters for Equation 3.4 that have the highest likelihood of fitting the observed data.

The likelihood function associated with the above distribution is:

$$\begin{aligned} L(\theta) = p(\vec{y}|X;\theta) &= \prod_{i=1}^{m} p(y^{(i)}|x^{(i)};\theta) \\ &= \prod_{i=1}^{m} h_\theta(x^{(i)})^{y^{(i)}}[1-h_\theta(x^{(i)})]^{1-y^{(i)}} \end{aligned} \tag{3.5}$$

Furthermore, the expression of the log-likelihood can be determined as:

$$\begin{aligned} \ell(\theta) &= \log L(\theta) \\ &= \sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1+y^{(i)}) \log [1-h_\theta(x^{(i)})] \end{aligned} \tag{3.6}$$

Thirdly, the value of the parameters that maximise the above expression can be determined using nonlinear programming solvers. With the obtained parameters, the probability of break for every record is given by using Equation 3.1. A threshold in the range 0 to 1 can be set for the given probability to differentiate breaks from no breaks.

3.2.3 Model fitting

Feature selection

A vector of features x is required for every record in order to fit the logistic regression model. The features are obtained by the attributes given in the dataset. However, if the features are utilised as they are without interactions or polynomial terms, valuable information could be lost. For example, it is known that age does not correlate linearly with the occurrence of pipe breaks (Andreou et al., 1987). Three stages occur in the life of pipelines (Kleiner & Rajani, 2001; Røstum, 2000; Gangl et al., 2007). In the first stage, a high number of breaks appears due to installation practices and material

imperfections. The break rate is then reduced during the second stage. The third stage includes accelerated pipe deterioration and failures. This non-linear behaviour could be captured and modelled with logistic regression by adding polynomial terms.

The addition of extra terms must be done systematically. Adding all computationally possible polynomial terms could result in overfitting for the model and may also add irrelevant variables, which may cause statistical problems such as multicollinearity. The selection of the interactions and higher order terms to include could be manually performed, based on engineering knowledge. However, this would require several manual iterations to obtain the best model given the data. Instead, the following procedure was implemented: (1). the maximum polynomial order that avoids overfitting was explored by iteratively increasing the order of the features and comparing the results of training and k-fold cross-validation; (2). a common feature selection algorithm was used to reduce the possible excess of features of the resulting model; and, (3). the behaviour of the variables was explored by a sensitivity analysis. If every augmented feature behaves in accordance with the available engineering knowledge, the model is accepted. For example, a decrease in pipe diameter increases break probability. Otherwise, the order of the feature is changed or the feature is removed, and the model fitted again.

The stepwise regression algorithm (The MathWorks Inc, 2024) was selected for the feature reduction analysis. This is a commonly used algorithm (Silhavy et al., 2017; Thomas & Peethamparan, 2017; Chester et al., 2018) that iterates from a base model and adds or removes terms given their impact on the model's performance to a selected criterion. Two common criteria used are the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The AIC aims to find the best model that approximates to the true model, (assumed not to be within the candidates set), and may tend to find complex models (Vrieze, 2012). The BIC aims to find the true model within the subset of models analysed and tend to generate simple models (Vrieze, 2012). If the true model is indeed within the subset of models the BIC selects it (Kuha, 2004). Contrary, if the true model is not within the subset of models, the AIC has a better performance (Kuha, 2004). The AIC was preferred for this study because the available data lacks important

variables, such as temperature and pressure transients. Therefore, it was reasonable to assume that the true underlying model cannot be generated or simplified with the data using the BIC approach.

ROC, PR curves and threshold selection

The use of the stepwise selection algorithm and the AIC helps in the identification of an appropriate subset of features. Nevertheless, the AIC does not give interpretation of the predictive power of the selected model. The area under the curve (AUC) of the receiver operating characteristics (ROC) curve is frequently used to assess the predictive capacity classification models. This metric, however, has been shown to be biased when there is a class imbalance in the data (Saito & Rehmsmeier, 2015). For example, in pipe breaks datasets, no breaks outnumber breaks. To address this issue, the AUC of the precision recall curve is utilised (Saito & Rehmsmeier, 2015). In this study, both metrics were applied for comparison. Derivation and explanation of both curves are provided as follows:

Both curves are derived from the confusion matrix for various thresholds (Haixiang et al., 2017). The confusion matrix includes four possible outcomes from a binary classifier within a two by two array. These are true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Three evaluation measures from the confusion matrix are of interest to develop the ROC and PR curves:

$$TPR = \frac{TP}{(TP + FN)} \tag{3.7}$$

$$FPR = \frac{FP}{(TN + FP)} \tag{3.8}$$

$$PREC = \frac{TP}{(TP + FP)} \tag{3.9}$$

Where TPR = True Positive Rate or Recall, FPR = False Positive Rate, $PREC$ = Precision. The ROC curve is derived by plotting the FPR in the x-axis and the TPR

in the y-axis for every possible threshold when evaluating the model. If the AUC of the ROC is greater than the area under a baseline $y = x$ (equivalent to 0.5), then the model has predictive power, and the closer the AUC to 1 the better it is. The ROC is a good measure when both classes, positive and negatives (1 and 0), are relatively balanced. When working with unbalanced data, the PR curve is recommended (Saito & Rehmsmeier, 2015). The PR curve is built by plotting the TPR or recall in the x-axis and the precision on the y-axis. A minimum threshold of $y = P/(P + N)$ is set, where P is the total positive examples in the data set and $(P + N)$ corresponds to the full data. If the AUC of the PR is greater than the AUC of the baseline, then the model has predictive power and the closer the AUC to 1 the better. This measure is a robust method to evaluate imbalanced datasets (Saito & Rehmsmeier, 2015).

ROC and PR curves evaluate the overall performance of a model under different thresholds. When it is desired to obtain specific predictions from the model, a value for the threshold must be set. In this case, a value for the threshold that maximises the F1-score generated by the model was determined. F1-score is a measure of the results performance of a model when there is an imbalance in the dataset. The F1 score is defined as follows:

$$F1 = \frac{2TP}{(2TP + FP + FN)} \tag{3.10}$$

K-fold cross-validation

To ensure that the results of the fitted model can be generalised, a cross-validation technique should be implemented. For this study k-fold cross-validation was chosen, whose formulation is as follows. The data S is randomly split into k disjoint subsets of m/k, where m is the number of records in S, and k is the number of desired subsets. In this case, $k = 5$ was chosen. Then, the model is evaluated as follows (Ng, 2007*a*):

for $j = 1, ..., k$

the model is trained on $S_1 \cup ... \cup S_{j-1} \cup S_{j+1} \cup ... \cup S_k$ to obtain the model h_j

Then, h_j is tested on S_j, and their associated ROC and PR curves are calculated.

Generalised ROC and PR curves for the model are finally estimated as the average of the ROC and PR curves for all h_j.

3.2.4 Sensitivity analysis

The main objective of this study is to develop and validate a method to assess the impact of controlling particular pressure components (metrics) on reducing pipe breaks. Previous studies have indicated that pressure reduction could potentially provide a decrease in pipe breaks, but quantification of the real impacts has not been determined yet. For this study, a machine learning algorithm was applied to predict pipe breaks in a water distribution network with the use of several covariates. After the proposed algorithm was validated for physical reliability and predictive capacity, then a procedure to assess how the number of breaks are affected by changes in pressure components was investigated. This procedure included a sensitivity analysis as follows:

The values of the mean pressure and pressure range in the network were varied by constant step reductions; i.e. the mean pressure obtained from the hydraulic model was iteratively reduced by steps of 5 m pressure head, and pipe breaks probabilities were recalculated for every step and stored for comparison. The results of the sensitivity analysis were then visualised in box-plots showing the change in the pipe breaks probabilities on the model results. Two points need to be noted. Firstly, if a threshold of 15 m pressure head was reached in the case of reducing the mean pressure, reduction of the pressure in that pipe was stopped. Secondly, for the case of pressure range a threshold of 0.01 m was set to stop reducing pressure.

3.3 Case study

3.3.1 Dataset summary

This study was built on data provided by the water utility of a medium-sized city in the UK. One dataset contained descriptive characteristics of the pipes in the network. This included, id, length, diameter, installation year, material and classification of the surrounding soil in terms of ferrous corrosivity and fracture potential. Another dataset contained recorded pipe repairs, which occurred in the network, and, in some cases, the type of pipe break was also provided. The period of the pipe breaks in the dataset encompasses from December 1994 to early June 2019. The development of statistics in the exploratory analysis was carried out with pipe breaks between 1995 and 2018. All available information was used for the modelling stage.

Maximum, minimum and mean pressure values for every node in the network were estimated from a hydraulic model used by the water utility. Average daily pressure values were derived for every pipe and associated in the dataset. General descriptive characteristics of the data are summarised in Table 3.1.

According to the dataset, three materials are predominantly used for the pipes installed in this water distribution network; cast iron (CI), plastic and asbestos cement (AC). However, the plastic pipes, although being the second largest group, have considerably fewer records of breaks than CI and AC pipes (Table 3.1). This may be related to the young age of these pipes that does not extend beyond 40 years. Based on the previous antecedents, the analysis was focused on cast iron and asbestos cement pipes.

3.3.2 Exploratory analysis

A section of the water distribution network, which was used in this study, is shown with the various pipe materials in Figure 3.1. The central area has the highest density of pipes with a considerable number of cast iron pipes. Asbestos cement and polyethene pipes

Table 3.1: Case study. From: Jara-Arriagada & Stoianov (2021).

Description	Value
Network length	∼6,800 km
Main materials (% of network length)	
Cast Iron	45%
Polyethene	22.3%
Asbestos Cement	19%
Diameters (% of network length)	
<= 100	43.7%
(100-150]	29.9%
(150-250]	14.0%
>250	12.4%
Burst data period for exploratory analysis	1995-2018
Number of breaks	18,767
Number of pipes affected by breaks	12,396
Percentage of the pipes affected by breaks per material	
Cast Iron	69.3%
Polyethene	4.6%
Asbestos Cement	14.9%

have been used as replacements for old cast iron pipes, and these can be found in specific sub-areas of the network. The oldest pipes in the network correspond to cast iron pipes with installation dates between 1840 and 1960.

The provided pressure data are from a steady state hydraulic model, which is used for engineering design and strategic planning by the water utility. No extended period simulation data, or actually measured pressure data, were used for this study. As a result, the pressure values are approximations for the mean, maximum and minimum pressures, which the network could experience over a period of time. Figure 3.2 (a) shows the distribution of the mean pressure for the cast iron pipes. The bars in the plots are divided by colours that represent the proportion of the diameter range for each pressure band. Similarly, Figure 3.2 (b) shows the distribution of the pressure range for the cast iron pipes. The distribution of the mean pressure over the total pipe length can be considered symmetric with a peak of around 35 m to 45 m pressure head. The spread of the mean pressure ranges from around 25 m pressure head to few cases with over 75 m pressure head. The distribution of the pressure range is right skewed. There is little variability in the pressure range. The majority of the pipes are in the range of 0 m to 10 m pressure

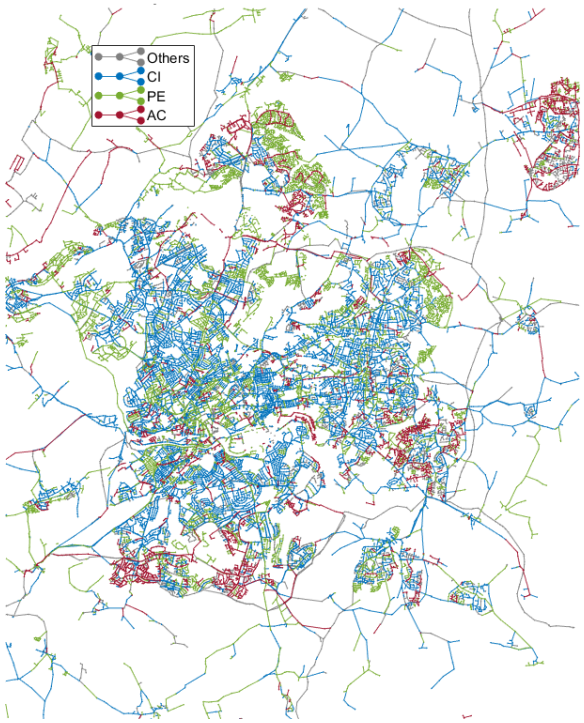


Figure 3.1: Central zone of the water distribution network. From: Jara-Arriagada & Stoianov (2021).

range with a high number of pipes in the range of 0 m to 2.5 m. As stated, this pressure range is estimated from the hydraulic model and does not include the actual dynamic hydraulic conditions.

In the analysed pipe breaks datasets, the attribute length represents the length of a pipe segment (id) based on arbitrary considerations relevant for building a hydraulic model from GIS. Therefore, the assigned length attribute has little relevance for the occurrence of a pipe break. However, it is an attribute of increasing exposition and probability to breaks (e.g. a larger number of pipe breaks occur in longer pipes). In Figure 3.3, the x-axis represents the resulting ranges of pipe lengths from grouping all pipes in the network in 10 groups with equal number of pipes (approximately 17,000 pipes in each bin). This analysis indicates that most pipe records are of pipes with short lengths and almost no breaks. Less than 5% of the total bursts have occurred in pipes with a length of up to 16.7 m, and these pipes constitute around 60% of all pipes for the considered case study. Only pipes with length greater than 16.7 m have been used in the presented model. This

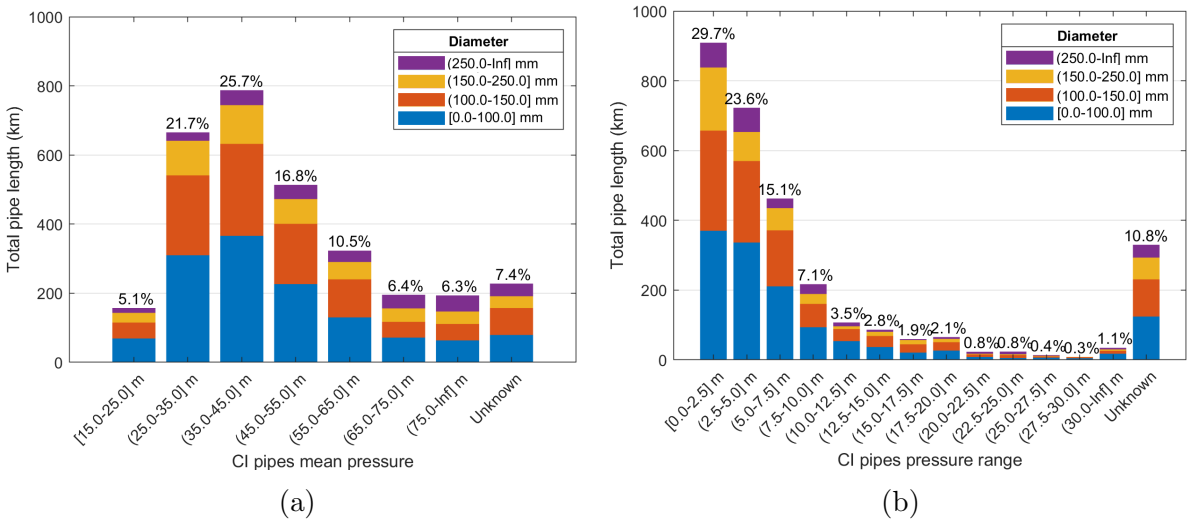


Figure 3.2: Distribution of pressure metrics for CI pipes. a) Mean pressure. b) Pressure range. From: Jara-Arriagada & Stoianov (2021).

approach reduces the number of excess zeros in the data without discarding pipes with break records.

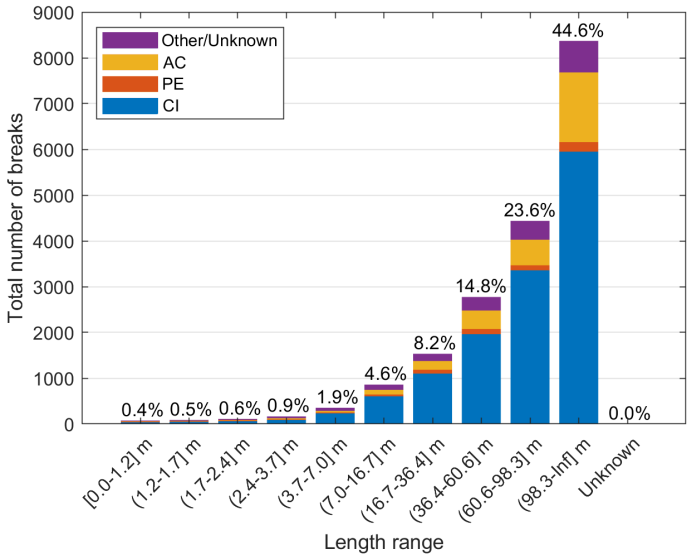
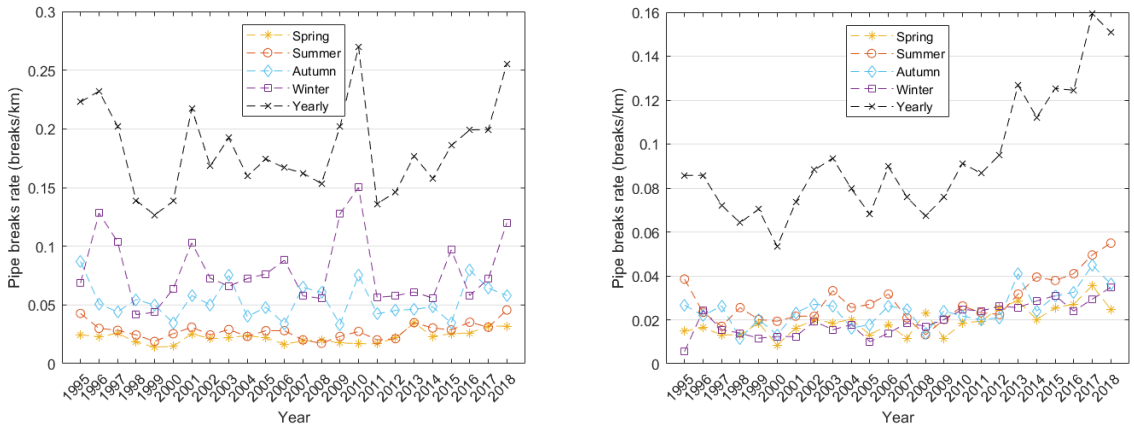


Figure 3.3: Number of breaks by pipe length range. The x-axis represents the resulting length ranges from grouping all the pipes in the network in 10 groups with equal number of pipes (approximately 17,000 pipes in each bin). From: Jara-Arriagada & Stoianov (2021).

The types of pipe breaks were provided for half of the records in the studied dataset. For the pipes with provided information about the break type, 30% described a circumferential break and the remaining 20% included ferrule, hole, joint, and longitudinal breaks. The large fraction of circumferential breaks suggested that pressure alone might not be the

main cause of breaks in this network. Circumferential breaks may result from bending stresses produced by thermal contraction (low temperature in the pipe and surroundings), soil differential movements, large cavities surrounding the pipe, inappropriate bedding practices or third parties. Longitudinal breaks, on the other hand, are associated to transverse stresses. These can be the result of the internal pressure in the pipe, and externals loads, such as cover loads, traffic, and soil expansion due to frosts (Rajani & Kleiner, 2001).

In the dataset, the failure (pipe breaks) rate for cast iron pipes (Figure 3.4 (a)) showed seasonality with the highest failure rate in winter months. The failure rate for asbestos cement pipes did not seem to be affected by seasonality (Figure 3.4 (b)), but showed an increasing trend over time. Polythene pipes exhibited a low number of failures, which was not enough to provide a consistent failure rate trend, and thus their graph was omitted. The failure rate difference for cast iron pipes in winter-autumn and summer-spring seasons led to the creation of two cohorts, namely a winter-autumn and a summer-spring cast iron break cohorts.



(a) (b)

Figure 3.4: Seasonal break rates. a) Cast iron pipes. b) Asbestos cement pipes. From: Jara-Arriagada & Stoianov (2021).

Table 3.2: Pipe cohorts for modelling. From: Jara-Arriagada & Stoianov (2021).

	Asbestos Cement	Cast Iron Winter	Cast Iron Summer
Network length (km)	∼552	∼1,736	∼1,478
Number of records	7,081	22,658	19,931
Number of breaks	947	4,731	2,004
Break % from the total	13%	21%	10%
Ratio break no break	1:6.5	1:3.8	1:8.9

3.3.3 Cohorts for analysis and model building

From the exploratory analysis, it was decided to implement three different pipe cohorts for modelling (Table 3.2): (1) asbestos cement cohort, including only asbestos cement pipes; (2) cast iron winter cohort, including only cast iron pipes with no breaks and breaks occurred in winter-autumn; and (3) cast iron summer cohort, which includes cast iron pipes with no breaks and cast iron pipes with breaks only in the summer-spring period. One break record per pipe was considered for modelling. In other words, if a pipe reported more than one break record the repeated break was not considered in the pipe cohort. Otherwise all the characteristic of the pipe except age would not change and this would not provide extra information for modelling within this particular approach. Details of the variables used for every subset are presented in Table 3.3.

Table 3.3: Variables used as input for modelling. From: Jara-Arriagada & Stoianov (2021).

Variable	Range		
	Asbestos cement	Cast iron winter	Cast iron summer
Length (m)	16.7 – 1,151	16.7 – 1265	16.7 – 1,265
Diameter (mm)	50 – 250	19 – 249	19 – 249
Pressure range (m)	0.03 – 56.21	0.03 – 94.7	0.03 – 94.73
Mean pressure (m)	11.51 – 132.04	8.99 – 133.3	8.99 – 132.8
Customer connections (n)	0 – 94	0 – 244	0 – 244
Age (years)	1 – 131	23 – 191	24 – 191

The dataset provided information for soil fracture potential and soil ferrous corrosivity, derived from an external source. These variables were of interest in modelling. However, after a series of iterations, it was found that including these variables resulted in counterintuitive results, such as decreasing pipe breaks probabilities with higher values of ferrous corrosivity. This behaviour was also observed by Boxall et al. (2007). To avoid

inconsistencies, it was decided to exclude these variables from the analysis.

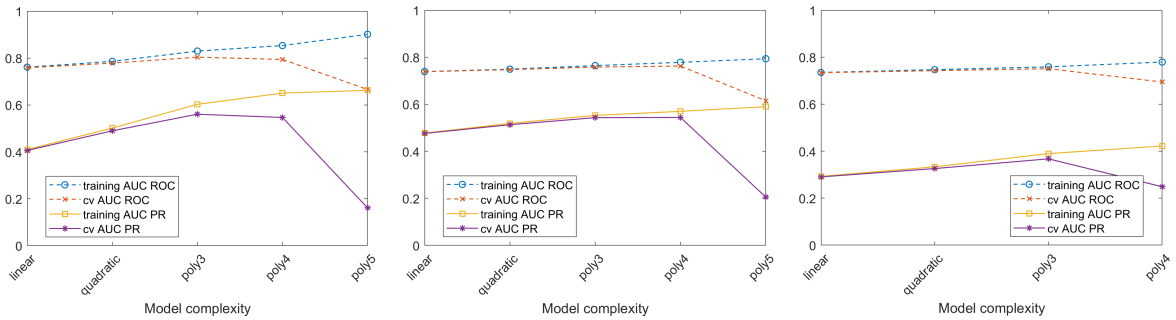
3.4 Results and discussion

This section presents and discusses results in two separate subsections. First, main results of the modelling approach and discussion of the selected model for every pipe cohort are presented. In the second subsection, analysis of the effect of pressure using the selected models, also for every pipe cohort, is discussed.

3.4.1 Main model results

Asbestos cement cohort

Exploration of the optimal polynomial degree for analysis showed that up to a 3rd degree polynomial for all variables, both training and validation provide similar results. Higher polynomial degrees show an overfitting pattern (Figure 3.5 (a)). Training results for both AUC ROC and PR curves increase as the polynomial degree increase. However, cross-validation results stagnate after increasing the model complexity to 4th degree polynomial and even decrease after 5th degree. This shows that the model stops generalising properly over 3rd degree polynomials by capturing noise in the data.



(a) (b) (c)

Figure 3.5: Areas under curve (AUC) of the receiver operating characteristics (ROC) and precision-recall (PR) curves for training and cross-validation (cv) by increasing model complexity. a) AC cohort. b) CI winter cohort. c) CI summer cohort. From: Jara-Arriagada & Stoianov (2021).

With the notion that up to a 3rd degree polynomial overfitting can be avoided, the stepwise regression algorithm was implemented to reduce the number of features. The original reduced model proposed alone by the stepwise regression included a cubic term of the feature age. However, this did not provide realistic results. The model output probabilities of almost zero for all pipes when changing their ages over 80 years old. To avoid the previous issue, the degree of the feature age was allowed to be expanded only up to a quadratic term with the stepwise regression. This change provided better consistency (similar results to Figure 3.6), however, the overall performance of the model decreased. The cross-validated AUC ROC decreased from 0.814 to 0.796 and the AUC PR decreased from of 0.581 to 0.526 (Table 3.4).

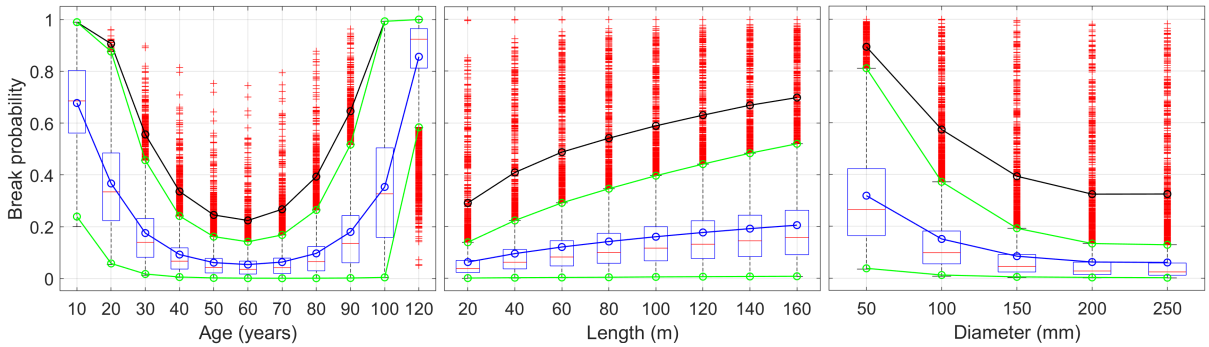


Figure 3.6: Coherence check for the asbestos cement model up to quadratic terms by varying values of age, length and diameter for all pipes. Similar trends were also obtained for both cast iron models. From: Jara-Arriagada & Stoianov (2021).

The decrease in performance by limiting the age variable reached similar values to a quadratic model (Figure 3.5 (a)). This prompted to compare the difference in performance of the previous model and a stepwise regression model up to a quadratic degree. The comparison showed that the quadratic model performed slightly worse in prediction in almost all aspects except recall (Table 3.4). However, this model used 17 terms in comparison with the 48 terms used by the model that explored cubic terms (except for the age feature). Also the probability trends estimated with the model by varying age, length and diameter were consistent to engineering knowledge (Figure 3.6). Despite the reduction in predictive power, it was decided that the simpler model up to quadratic terms would be more appropriate for inference and therefore selected for further analysis. Table 3.5

presents the resulting logistic regression models utilised for further analysis in every pipe cohort.

Cast iron winter cohort

The approach, which was applied for the AC cohort, showed that for the CI cohort with winter breaks no overfitting was observed up to 4th degree polynomial (Figure 3.5 (b)). However, the difference in performance between a quadratic and 4th degree model was less strong, and resulted in the selection of a model with interactions up to quadratic terms. The model outputs agreed with engineering knowledge. For example, the model estimates an increase in pipe breaks probabilities if the length of pipes was increased. In addition, the breaks probabilities increase if the diameter of pipes was decreased. An increase in both pressure metrics increases break probabilities. An increase in customer connections increases break probabilities. The age shows a consistent pattern with high break probability when pipes are recently installed and a decrease in break probability later in their operational life. However, the rise in break probability start to increase in general only after 140 years, which may be due to the quality of the records. The dataset contains pipes over 180 years old and many pipes over 100 years without any reported break that may influence the data fitting.

Table 3.4: Performance metrics by pipe cohorts and maximum polynomial degree explored by the stepwise regression using the Akaike Information Criterion (AIC). From: Jara-Arriagada & Stoianov (2021).

Cohort	AC cohort			CI winter cohort	CI summer cohort
	poly3	poly3 (age^2)	quadratic	quadratic	quadratic
cv AUC ROC	0.814	0.796	0.78	0.748	0.744
cv AUC PR	0.581	0.526	0.491	0.515	0.33
Threshold	0.22	0.32	0.24	0.26	0.17
cv F1-score	0.53	0.501	0.48	0.48	0.35
cv Recall	53.7%	42.3%	48.8%	54.7%	43.6%
cv Precision	49.5%	57.6%	46.4%	42.9%	29.2%
N° features	44	48	17	21	19
coherence	no	yes	yes	yes	yes

Table 3.5: Covariates and coefficients for the logistic regression models utilised for analysis in every pipe cohort. From: Jara-Arriagada & Stoianov (2021).

Cohort	Formula $[logit(p) = \theta^T x]$
Asbestos cement	$logit(p) = -2.77 - 0.92 \cdot A + 0.33 \cdot A^2 + 0.52 \cdot L - 0.7 \cdot D + 0.18 \cdot MP + 0.24 \cdot CC + 0.11 \cdot (MP \cdot A) + 0.15 \cdot PR - 0.05 \cdot CC^2 + 0.11 \cdot D^2 - 0.1 \cdot (D \cdot PR) - 0.11 \cdot (D \cdot A) - 0.07 \cdot (PR \cdot CC) + 0.07 \cdot (L \cdot CC) + 0.06 \cdot (L \cdot PR) - 0.05 \cdot (CC \cdot A)$
Cast iron winter	$logit(p) = -1.89 - 0.68 \cdot A + 0.29 \cdot A^2 + 0.43 \cdot L - 0.46 \cdot D + 0.31 \cdot CC + 0.16 \cdot MP - 0.08 \cdot (L \cdot CC) + 0.06 \cdot L^2 - 0.02 \cdot CC^2 - 0.04 \cdot MP^2 - 0.07 \cdot (L \cdot A) + 0.06 \cdot (CC \cdot A) + 0.07 \cdot (D \cdot A) + 0.05 \cdot (L \cdot D) + 0.04 \cdot D^2 - 0.05 \cdot (D \cdot CC) + 0.03 \cdot (L \cdot MP) + 0.04 \cdot (PR \cdot MP) - 0.02 \cdot PR^2 - 0.01 \cdot PR$
Cast iron summer	$logit(p) = -2.9 - 0.57 \cdot A + 0.27 \cdot A^2 + 0.46 \cdot L - 0.36 \cdot D + 0.37 \cdot CC + 0.09 \cdot L^2 + 0.17 \cdot MP - 0.03 \cdot CC^2 + 0.09 \cdot (D \cdot A) + 0.07 \cdot D^2 - 0.08 \cdot (L \cdot CC) + 0.04 \cdot (CC \cdot A) - 0.05 \cdot (PR \cdot A) - 0.03 \cdot (MP^2) + 0.05 \cdot (PR \cdot MP) + 0.05 \cdot (L \cdot PR) - 0.05 \cdot (PR \cdot CC) + 0.02 \cdot PR$

Where: L = length, D = diameter, A = age, PR = pressure range, MP = mean pressure, CC = customer connections

Cast iron summer cohort

Similarly to the other two cohorts, a model with interactions and quadratic terms provided a similar performance to higher order degrees (Figure 3.5 (c)). However, in this particular case, the precision of the model was considerable low. Although the model reached a cross-validation value of 0.744 for its AUC ROC which could be considered acceptable, it only achieved a value of 0.33 for its AUC PR. The value of the AUC PR curve is still far over the threshold that separates the model for a random classifier. For a threshold that optimises its F1-score, the model has a recall of 43.6% but a precision of 29.2%. The meaning of this is as follows: Over all the pipes that are given to the model, this is capable of identifying almost half of the pipes that actually break within its predictions. However, from all the models' break predictions, only around 29% are true. Considering that breaks are a minority class within a large population, this still indicates predictive power of the model, but that can be further improved. When the predictive power of the model is analysed to assess no breaks predictions, the model has a recall of 87.2% and a precision of 93.5%. This example can be extended to the other two cohorts.

Note that the results from this model need to be considered cautiously as these show higher signals of bias. The quality of the data and the features explored play an important role.

More factors affecting pipe breaks in summer may be missing, which may have led to the low precision and to miss important interactions; for example, the inclusion of pressure transients.

3.4.2 Sensitivity analysis

The models obtained for the AC cohort and both CI cohorts were used to simulate the effects of pressure control. The original values provided from the hydraulic model for every pipe were decreased in steps until the regulatory requirement in England and Wales for minimum pressure head of 15 m in a water distribution pipe for the mean pressure. The pressure range was allowed to decrease for every pipe until almost 0. The box plots of Figure 3.7 were developed to gain understanding of how the change in pressure affects the trends in the distribution of the predicted probabilities of the models. Tables 3.6 and 3.7 show the quantitative results of pressure reduction on the number of predicted breaks by the three models.

There is a reasonable number of outliers shown in the box plots of Figure 3.7. An explanation for this is the fact that most of the pipes used for modelling have not experienced pipe breaks (Table 3.2). Therefore, it is expected that most of the predictions given by the models will concentrate in low values. This lead to treat break predictions with high degree of confidence as outliers. The distance between the lower quartile and lower whisker represents 25% of the data with lowest break probability. It is expected that this would not have a high variation as pressure decreases since their break probability is already low. Aspects of interest are the distance between the upper whisker and upper quartile, and the spread of the outliers. Decrease of these ranges as pressure reduces indicates reduction of the break probability trend for the pipes with more chance to break. The thick black line indicates the mean position of the outliers while the blue line indicates the mean of the data. Depending of the optimal threshold selected to differentiate breaks from no breaks, slight changes may lead to a large reduction of predicted breaks. Consequently, Table 3.6 and Table 3.7 are provided to understand the actual quantitative changes in

pipe break predictions for the models as pressure metrics decrease.

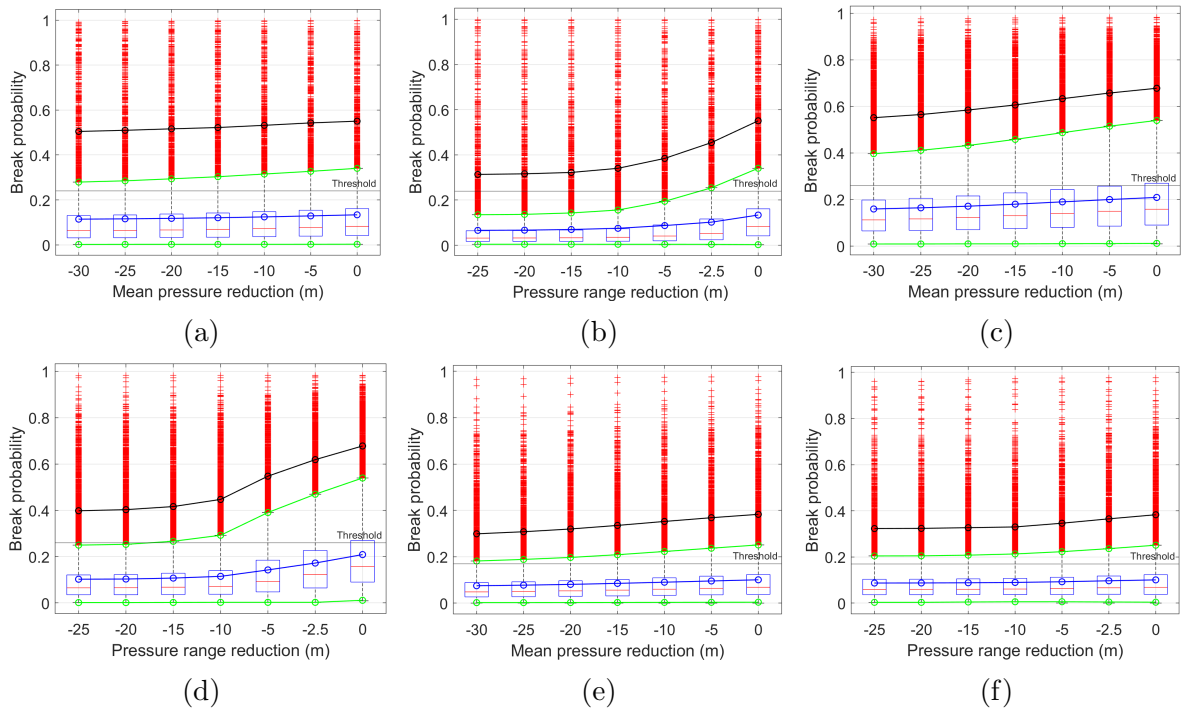


Figure 3.7: Pipe break probability change by decreasing pressure metrics (from right to left). The x-axis in the plots indicates the mH_2O reduction applied to every pipe in the network if possible. a) AC cohort mean pressure. b) AC cohort pressure range. c) CI winter cohort mean pressure. d) CI winter cohort pressure range. e) CI summer cohort mean pressure. f) CI summer cohort pressure range. From: Jara-Arriagada & Stoianov (2021).

Asbestos cement cohort

Results of the sensitivity analysis for the asbestos cement cohort indicate that both mean pressure and pressure range have potential effects on reducing pipe breaks. There is a slight and constant decrease of the upper whisker for the box plot that shows the predicted break probabilities as function of the mean pressure reduction (Figure 3.7 (a)). This led to a decrease of 10% in predicted breaks when applying a reduction of 10 m pressure head in every pipe. An 18% reduction was observed by decreasing the mean pressure in every pipe up to 20 m (Table 3.6).

The reduction in pressure range was of particular interest as it suggests an almost exponential decrease in pipe breaks up to 10 m pressure range reduction (Figure 3.7 (b)).

After 10 m pressure range reduction the reduction in pipe break probabilities flattens. This may be a reflection of the quality of the data used. As shown in Figure 3.2 the distribution of the pressure range is right skewed and with little variability for most of the pipes. A reduction in pressure range of 5 m pressure head leads to a decrease of 40% in pipe breaks as estimated by the model. A reduction in pressure range of 10 m pressure head leads to a high decrease of 54% in predicted pipe breaks (Table 3.7). These results suggest that reduction of pressure range may be an important factor to consider to reduce pipe breaks for asbestos cement pipes. These results however, should be confirmed with a highly calibrated model that truly reflect the pressure range distribution. However, the failure mechanism by which pressure range would be influencing pipe breaks does not seem intuitive. Pressure range, could be related in fact as a surrogate variable for areas with larger pressure fluctuations given the variability in the maximum and minimum pressure in the network.

Cast iron winter cohort

The model for cast iron pipes winter cohort indicates an even higher potential for reducing pipe breaks with reducing both mean pressure and pressure range. A constant decrease in pipe break probabilities is estimated by the model when reducing the mean pressure for every pipe (Figure 3.7 (c)). This translates to a 16% reduction in the number of pipe breaks predicted by the model by a mean pressure decrease of 10 m for every pipe. The number of predicted breaks decrease by 30% with a mean pressure reduction of 20 m. (Table 3.6).

The largest reduction in pipe breaks prediction is obtained by decreasing the pressure range. There is a steep decrease in the distance between the upper whisker and upper quartile as the pressure range reduces for every pipe. Also, the outliers tend to concentrate and approximate to the upper whisker (Figure 3.7 (d)). Similar to the asbestos cement, the reduction in break probabilities flattens beyond a decrease in 10 m pressure range, possibly due to the pressure range distribution. High reductions in breaks predictions by

43% and 61% are given by the model when decreasing the pressure range in 5 and 10 m respectively.

Results from this model are of particular interest because of two reasons: (1) This model has the best predictive power from the three pipe cohorts analysed (Table 3.4), and (2) it also makes use of most of the available data (Table 3.2). This increases the confidence of the obtained results and the notion that pressure range (or in fact pressure fluctuations) may play an important role on the occurrence of pipes breaks. Another important consideration is that the model only considered cast iron pipes breaks occurred in winter. Therefore, it may be possible that high pressure range or fluctuations causing fatigue might accentuate the effects of low temperature on cast iron pipes (e.g. the embrittlement of cast iron pipes at low temperatures combined with the impact of fatigue and pressure).

Cast iron summer cohort

Results of the sensitivity analysis indicate that mean pressure reduction also has positive impacts on this pipe cohort. There is a constant reduction of the upper whisker and the mean of the outliers as the mean pressure decrease (Figure 3.7 (e). A reduction of the mean pressure by 10 m and 20 m leads to 18% and 32% decrease in predicted breaks respectively (Table 3.4).

Contrary to the CI winter cohort, the model for the CI summer cohort shows a lower impact of the pressure range on the probability of pipe breaks (Figure 3.7 (f), Table 3.7). It is hypothesised is that low temperature embrittles the cast iron material, and this exacerbates the impact of pressure range or potentially fatigue on pipe breaks in winter. The model for CI summer cohort is the most inaccurate model in terms of predictive power from the three models analysed. Therefore, results may be biased. The trends showed by this model, however, are consistent with the estimates from the other two models.

Table 3.6: Effect of reducing the mean pressure in the model's predictions. From: Jara-Arriagada & Stoianov (2021).

Model	T	Number of predicted pipe breaks (% decrease)		
		Given pressure	10 m reduction	20 m reduction
AC cohort	0.24	1006 (0%)	910 (10%)	827 (18%)
CI winter cohort	0.26	6023 (0%)	5078 (16%)	4215 (30%)
CI summer cohort	0.17	2988 (0%)	2438 (18%)	2017 (32%)

T indicates the threshold used for classification that maximises the F1 score for the models

Table 3.7: Effect of reducing the pressure range in the model's predictions. From: Jara-Arriagada & Stoianov (2021).

Model	T	Number of predicted pipe breaks (% decrease)		
		Given pressure	5 m reduction	10 m reduction
AC cohort	0.24	1006 (0%)	603 (40%)	464 (54%)
CI winter cohort	0.26	6023 (0%)	3417 (43%)	2348 (61%)
CI summer cohort	0.17	2988 (0%)	2478 (17%)	2347 (21%)

T indicates the threshold used for classification that maximises the F1 score for the models

3.5 Conclusions

This chapter has utilised an extensive dataset of pipe breaks, and information of average steady state metrics of water pressure in order to confirm previous evidence of the influence of water pressure on pipe breaks. Particular attention was given to model pipe breaks using best practices in machine learning to ensure physical reliability of the obtained models, and to produce a sensitivity analysis of the magnitudes of pressure needed to produce relevant pipe break reductions. Logistic regression with polynomial terms was the modelling approach selected for this analysis.

The results indicate a 18% decrease in the number of pipe breaks when reducing the mean pressure up to 20 m for asbestos cement pipes. For the same mean pressure reduction, 30% and 32% reduction in pipe breaks were estimated for cast iron pipes in winter and summer respectively. A reduction in 10 m pressure range suggests a major 54% reduction in predicted pipe breaks for asbestos cement pipes. Also, the same pressure range reduction suggests a major 61% reduction in pipe breaks for cast iron pipes in winter. This effect, however, was less pronounced for pipe breaks for cast iron pipes in summer months as the model suggests a reduction of 21% in pipe break predictions. The results, however, are

influenced by the hydraulic pressure data, which was obtained from a hydraulic model. Actual pressure measurements with high spatial and temporal resolution, or a calibrated hydraulic model would be particularly beneficial to complement and expand the presented analysis.

In general, the results show a positive correlation of mean pressure and pressure range to pipe breaks. These results suggest that pressure control has the potential to enhance the reliability of water distribution networks by reducing pipe breaks through a decrease of average pressure metrics. However, although the magnitudes needed for reduction of mean pressure are align with industry experience, it remains challenging to mechanically justify the impact of mean pressure reduction alone. Previous research linked water pressure to pipe breaks but lacked a discussion of the magnitude of mean pressure and the physical mechanism involved in the failure of distribution pipes, which are affected by lower pressure levels than transmission mains.

As predicted by the developed models and observed through the exploratory analysis, the mean pressure in the pipes is insufficient to directly cause pipe breaks, even in severely deteriorated pipes. A similar conclusion can be drawn regarding pressure range, whose effect is more pronounced than that of mean pressure. Nevertheless, the relationships identified in both prior research and this study, which leverages a large dataset, show that even these relatively low levels of mean pressure and pressure range influence pipe breaks.

In line with the cumulative pressure-induced stress hypothesis, it is reasonable to propose that the observed relationships are linked to pressure fluctuations and dynamic pressure changes within the network. Areas experiencing a wide pressure range may be particularly vulnerable due to the influence of control devices and user activity, which can intensify pressure fluctuations, thereby contributing to long-term fatigue damage. This could explain the significant impact of this variable in the developed models. Furthermore, as will be discussed in the next chapter, mean pressure influences the rate of fatigue crack propagation due to pressure cycling, which in turn contributes to the long-term fatigue

affecting pipes.

Overall, this chapter revisited the influence of average metrics of water pressure on pipe breaks as reported in past research. The results of this chapter showed specific trends and magnitudes of the reduction of pipe breaks through reduction of average pressure metrics, complementing past research. However, although the results align with past literature, they open the motivation to investigate the impact of pressure dynamics and their relation with pipe breaks, as presented in the next chapter.

Chapter 4

Fracture mechanics of pressure-induced fatigue failures

Related publication: Jara-Arriagada, C, & Stoianov, I. (2024) Pressure-induced fatigue failures in cast iron water supply pipes. *Engineering Failure Analysis*. 155, 107731. doi:10.1016/j.engfailanal.2023.107731

4.1 Introduction

The previous chapter has shown how average metrics of water pressure relate to pipe breaks in water distribution networks (WDNs). However, the magnitude of the effects, particularly in areas with higher pressure variation, suggests that static pressure alone doesn't fully explain the observed statistical impact. A possible explanation for this problem is the hypothesis and concept of cumulative pressure-induced stress (CPIS), which was introduced in the Literature Review (Chapter 2). This concept has been initially proposed in Hoskins & Stoianov (2017), and further developed by Rezaei (2017) offering a potential framework for understanding pipe breaks. However, CPIS remains in early stages of research, requiring further investigation, especially in terms of how pressure transients contribute to pipe breaks. This chapter aims to fill this gap, by presenting a

comprehensive theoretical review and analysis of the processes involved in the fatigue failure of pipes, and the conditions needed for pressure transients to be relevant to the fatigue of cast iron pipes. These pipes were selected given the large majority of ageing pipes of this material in UK water networks, which requires significant attention.

The analysis in this chapter adopts a systematic approach, considering the properties of pipes in the wear-out phase, along with hydraulic and loading conditions that account for both pressure fluctuations and mean operational pressure. This chapter addresses the following research questions: (1) What types of deterioration are observed in cast iron pipes; (2) What magnitude and frequency of pressure fluctuations can lead to pressure-induced fatigue failures; (3) To what extent mean pressure does influence pressure-induced fatigue failures; (4) Should pressure reduction schemes aim to reduce mean pressure, pressure fluctuations or both in order to reduce the occurrence of pipe breaks in ageing cast iron pipes.

To answer these questions, a thorough review is conducted on cast iron pipes, their material properties, the effects of corrosion-related deterioration, the observed failures in cast iron pipes, and the theory behind fatigue failure. Based on this review, a theoretical analysis is then carried out to evaluate the impact of pressure transients on fatigue, supporting the concept of cumulative pressure-induced stress.

4.2 Cast iron pipes in water distribution networks

This section presents key factors and theory for the application of fracture mechanics in cast iron pipes. In particular, manufacturing processes, microstructure, effect of corrosion on pipe wall thickness, and fatigue experiments are discussed. These were then included in the application of the fracture mechanics and fatigue crack propagation theory presented in section 4.3.

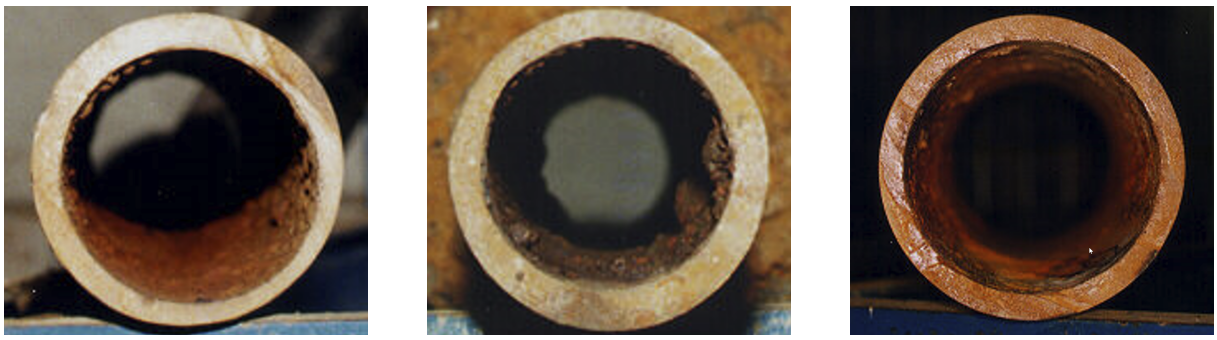
Cast iron is in essence an iron and carbon alloy with additions of silicon, manganese and other elements in minor quantities included to improve the casting process and its

corrosion resistance (Cast Iron Pipe Research Association, 1952). The carbon content usually varies between 2.5% to 4%. This content of carbon significantly reduces the melting point of cast iron in comparison to steel, thus providing greater fluidity to cast complex shapes (Makar & Rajani, 2000). The carbon content in cast iron pipes is usually observed as graphite flakes, spread within the mass of metal. These graphite flakes control most of the mechanical properties of cast iron (Cast Iron Pipe Research Association, 1952; Baker, 1978). The metal part of cast iron can be either ferrite, pearlite or a mixture of both (Makar & Rajani, 2000). An extensive body of research has studied the material properties of cast iron (Speidel, 1981; Romaniv et al., 1990; Hornbogen, 1985; Baker, 1978; Conlin, 1991; Makar & Mcdonald, 2007). The following sub-sections summarise the metallurgy of cast iron and its mechanical properties, which are utilised in the mechanical analyses presented in section 4.4.

4.2.1 Manufacturing processes of cast iron and the effect on pipe wall thickness

Four distinct manufacturing processes were developed and implemented for cast iron pipes. These processes are horizontal casting, vertical or pit casting, centrifugal or spun casting in sand mould, and centrifugal or spun casting in metal moulds (Cast Iron Pipe Research Association, 1952). All cast iron pipes were horizontally cast until 1845, which included the pouring of the melted iron in a horizontal mould. This procedure caused an uneven distribution of the metal in the mould leading to large variability and non-uniformity in circumferential pipe sections (Figure 4.1a) and wall thickness. Cast vertically in a pit was introduced in around 1845, in which the moulds were vertically placed. However, variability and non-uniformity in pipe wall thickness continued to be observed. In 1918, centrifugal or spun casting was introduced. This manufacturing process resulted in a more uniform pipe wall thickness (Figure 4.1c).

The adoption and use of new manufacturing processes for cast iron pipes were not governed by appropriate standards for a prolong time period, during which cast iron pipes were



(a) (b) (c)

Figure 4.1: Variability and non-uniformity in pipe wall thickness due to manufacturing practices. a) & b) Large non-uniformity in pipe wall thickness, potentially from horizontal or vertically pit casting process. c) Spun cast iron pipe. Photos provided by Bristol Water, UK. From: Jara-Arriagada & Stoianov (2024).

extensively installed. For example, British standards for pit cast iron and spun cast iron manufacturing processes were introduced in 1917 (British Standards Institute, 1917) and 1945 (British Standards Institute, 1958) respectively. Consequently, there were significant variations in the manufacturing processes and quality controls for a large stock of cast iron pipes, which are still in use.

4.2.2 Microstructure of cast iron pipes

Manufacturing processes also influenced variations in the metal composition and the geometry of the graphite flakes in cast iron pipes (Makar & Rajani, 2000). These graphite flakes play a major role in defining the mechanical properties of cast iron pipes; and in particular, the tensile strength, fracture toughness, and the stress intensity threshold for fatigue (Baker, 1978). The form, size, and distribution of the graphite flakes are defined in the ASTM A247 (ASTM, 2019).

In terms of the type of graphite flakes, there are no major differences resulting from the manufacturing processes for cast iron pipes. The four types of manufacturing processes produce type VII graphite flakes according to ASTM A247 (ASTM, 2019). However, the matrix (size and distribution) of the graphite flakes, as well as the metal part largely vary between static and centrifugal castings. These variations result from the different cooling

rates. Large flakes and ferrite are produced by slow cooling rates, while a faster cooling rate tends to produce smaller graphite flakes and pearlite. A quick cooling rate leads to fine flakes and ferrite (Makar & Rajani, 2000).

Horizontal and vertical cast iron pipes require a longer cooling time than spun cast iron pipes. Consequently, it is usual to observe graphite distribution of types A, B, and C in horizontal and pit cast iron pipes, while spun cast iron pipes are characterised by type D distribution of graphite flakes (Makar & Rajani, 2000). In terms of the metallic grains, pit cast iron pipes exhibit predominantly pearlite grains, while spun cast iron pipes are predominantly ferrite. Grain sizes in pit cast iron pipes are also much larger than in spun cast iron pipes due to the slow cooling process, which allows more time for individual grains to form (Makar & Rajani, 2000).

4.2.3 Graphite flakes as crack initiators

The size of the carbon flakes significantly affects the mechanical properties of cast iron pipes. Graphite flake sizes are classified in eight descending categories from class 1 to 8. For the case of pit cast iron pipes, the graphite flakes can reach sizes of the class 3, 2 and 1. The usual range is between 0.3 mm to 0.6 mm. However in some instances, graphite flakes can reach up to 1 mm to 2 mm (Baker, 1978). Much smaller graphite flakes are seen in spun cast iron pipes (< 20 μm). This difference in graphite flakes has been identified as the cause for large differences in the tensile strength and fracture toughness between pit and spun cast iron pipes (Makar & Rajani, 2000). The larger carbon flakes in pit cast iron makes it easier for cracks to initiate and propagate (Makar & Rajani, 2000). Table 4.1 presents a summary of common values reported in the literature for the main mechanical properties for cast iron pipes.

In summary, pit cast iron pipes are more likely to develop large cracks as the size of the graphite flakes produced in these pipes are larger than in spun cast iron pipes. Nevertheless, because of the variable and non-uniform pipe wall thickness, greater safety factors were used for pit cast iron pipes, and therefore pipes with thicker walls than required were

Table 4.1: Reported measured mechanical properties for cast iron (pit and spun). From: Jara-Arriagada & Stoianov (2024).

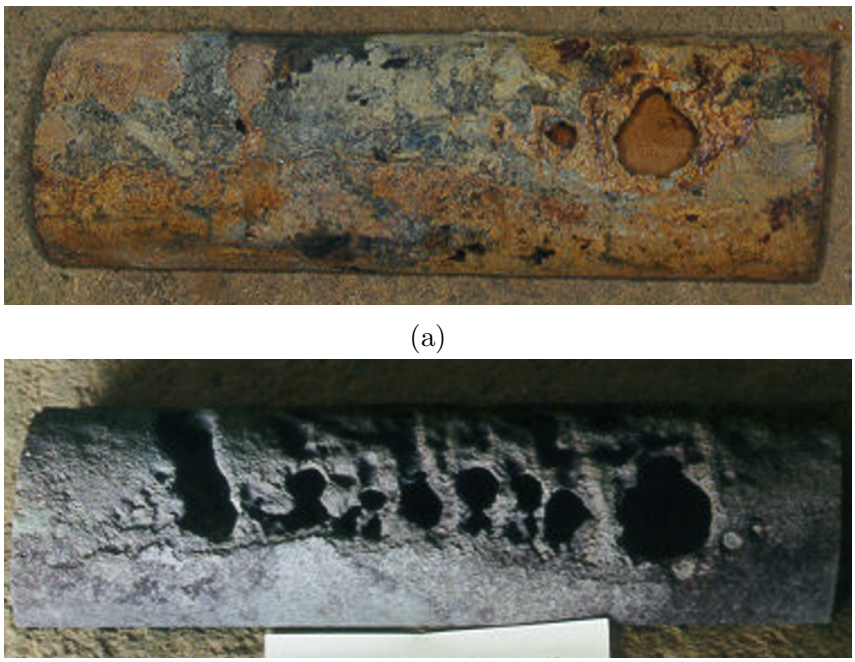
Material property	Pit cast iron	Spun cast iron	Observation*	References
Tensile strength [MPa]	68 - 156	157 - 305		(Makar & Rajani, 2000), (Cullin et al., 2015) (Debnath et al., 2021)
Fracture toughness [MPa$\sqrt{m}$]	7.9 - 24.3	11.9 - 20.36		(Makar & Rajani, 2000), (Cullin et al., 2015) (Debnath et al., 2021)
Stress intensity threshold [MPa$\sqrt{m}$]	2	-	R = 0.7	(Bulloch, 1995)
	5.7	-	R = 0.5	(Bulloch, 1995)
	8	-	R = 0.1	Mohebbi et al. (2010)
Paris constant C [m/cycle]	1.806e-12	-	R = 0.7	(Bulloch, 1995)
	1.213e-14	-	R = 0.5	(Bulloch, 1995)
	6.559e-15	-	R = 0.1	(Mohebbi et al., 2010)
Paris constant m [-]	7.255	-	R = 0.7	(Bulloch, 1995)
	7.647	-	R = 0.5	(Bulloch, 1995)
	7.379	-	R = 0.1	(Mohebbi et al., 2010)

* R is the ratio $\frac{\text{pressure}_{\text{min}}}{\text{pressure}_{\text{max}}}$ at which the fatigue experiments were performed.

produced. In contrast, thinner pipe walls were used in spun cast iron pipes because of the uniformity of the pipe wall thickness, and the larger tensile strength of the material. These thinner pipe wall thicknesses, however, makes spun cast iron pipes more susceptible to pit corrosion damage, which is discussed in the next section.

4.2.4 The effects of corrosion on pipe wall thickness, pits and cracks

Corrosion describes processes involving electrochemical reactions, in which a refined metal is converted to a more stable form such as an oxide, hydroxide or sulphide (Anderson, 2017). During these processes metal is lost, and therefore, the structural resistance of the material is degraded (Romanoff, 1957). The most common type of corrosion process observed in cast iron pipes is pitting corrosion with graphitisation (Petersen & Melchers, 2012). The graphitisation process includes the leaching of the metallic components from the cast iron, and the accumulation of graphite on top of the corrosive layer when a pit is formed in the metal. This layer of graphite provides the metal with the appearance of a non-corroded surface (Petersen & Melchers, 2012). However, the resistance of the material filling the pit is of much lower strength than the cast iron, making it prone to failure at



(a)

(b)

Figure 4.2: The extent of graphitic corrosion in a cast iron pipe as identified by sandblasting. a) Pipe sample before sandblasting. b) Pipe sample after sandblasting. Samples provided by Bristol Water, UK. From: Jara-Arriagada & Stoianov (2024).

low loadings (Petersen & Melchers, 2012). Usually, a corrosion pit due to graphitisation can only be identified by sandblasting a pipe (Figure 4.2). In addition, corrosion pits can occur almost everywhere in a pipe (Li & Mahmoodian, 2013), and because corrosion processes are affected by various factors, precise locations of corrosion spots are difficult to predict (Petersen & Melchers, 2012).

There are various phases in the corrosion process for cast iron pipes (Petersen & Melchers, 2012; Soltani Asadi & Melchers, 2017*a*). For this investigation, the last phase, in which the corrosion process reaches a steady state, is of particular interest to assess the long term impact of corrosion at the onset of and during the wear-out phase (Petersen & Melchers, 2012). Published research has suggested that in the long term (+30 years installation of a cast iron pipe), pitting corrosion can reach a steady state (Soltani Asadi & Melchers, 2017*b*, 2018; Kim et al., 2007). The suggestion comes from evidence that maximum observed corrosion pit depths in cast iron pipes do not exceed 10 to 12 mm in very aged pipes. Also for most pipes, it seems that 7 mm is a common minimum corrosion pit depth independently of pipe age (Soltani Asadi & Melchers, 2017*b*, 2018).

Soltani Asadi & Melchers (2018) have suggested that the build-up of material, which is generated during pitting corrosion tend to isolate the corrosion interface from the exterior and thus reducing its rate of vertical propagation during longer exposure periods. As a conclusion, it is reasonable to assume a minimum value of 6-7 mm for existing corrosion pits or amalgamated corrosion patches in cast iron pipes independently of their age.

Typical maximum values for corrosion depths (10 mm to 12 mm) can also be compared against the proposed wall thicknesses for cast iron pipes given by British Standards (i.e. BS 78:1917 and BS 1211:1945). Figure 4.3 shows design pipe wall thicknesses for different diameters and classes of cast iron pipes. These thicknesses are visually compared to the limits of 10 mm and 12 mm for maximum observed corrosion depths whose areas of influence are marked in colour. From Figure 4.3, it can be observed that small diameter pipes seem at most risk to be affected from wall thickness reduction because of their reduced thickness, which falls within maximum corrosion pit depths. This observation is reinforced by the analysis of historic pipe breaks, which shows that small diameter cast iron pipes have much higher failure rates (Figure 3.6). A similar conclusion can be drawn from the wall thickness difference between pit and spun cast iron pipes (Figure 4.3b). The reduced wall thickness of spun cast iron pipes makes these pipes more vulnerable to corrosion.

In addition to the pipe wall thickness reduction that corrosion generates on pipes, it has been recently investigated that corrosive soil environments, especially with high moisture contents and pH of around 5 can produce a reduction on the nanomechanical properties of cast iron (Wasim et al., 2021). This reduction in nanomechanical properties affects in turn bulk properties such as elastic modulus, tensile strength and fracture toughness (Wasim et al., 2021). In this research, for simplicity, and as a conservative approach, it is considered that corrosion only reduces pipe wall thickness, and does not affect the mechanical properties of cast iron pipes.

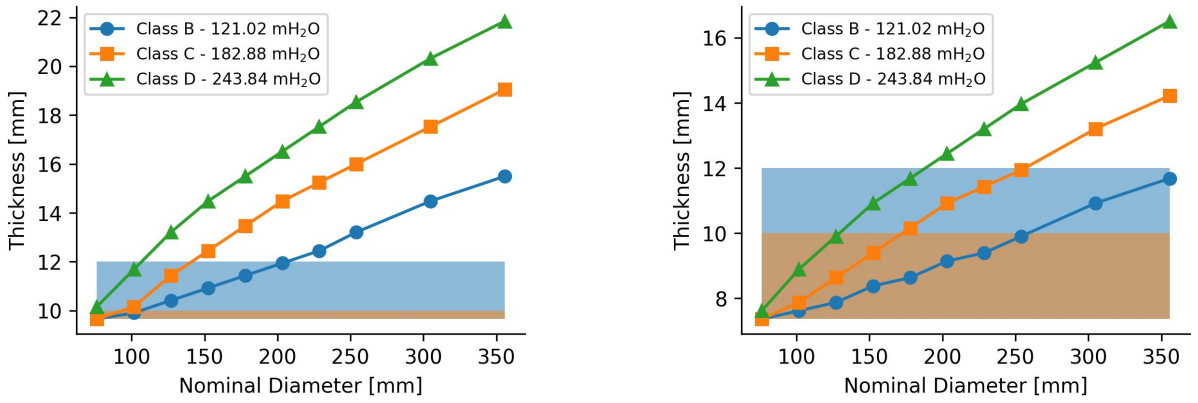


Figure 4.3: Pipe wall thicknesses for pit and spun cast iron pipes according to British Standards BS-78 and BS-1211, and comparison to maximum observed corrosion pit depths (between 10 and 12 mm). a) Pit cast iron pipe thicknesses. b) Spun cast iron pipe thicknesses. From: Jara-Arriagada & Stoianov (2024).

4.2.5 Fatigue experiments on cast iron pipes

This section reviews and presents research on fatigue-induced failures in cast iron pipes. This review together with key conclusions drawn in the previous sections are used for the mechanical analysis presented in Section 4.4.

In general, fatigue of materials can be analysed by the use of S-N curves, ϵ-N curves, and fatigue crack propagation experiments. The first two approaches deal with the total fatigue life of a specimen given their initial conditions. The results of such approaches are only valid to explain the failure of a material under the original conditions at which it was tested. In contrast, fatigue crack propagation experiments offer major generalisation as the results from testing can be used by any type and size of cracks. Given the different states of deterioration and the presence of cracks in buried cast iron pipes, fracture mechanics is more appropriate to perform pressure-induced fatigue analysis.

Published research describes various studies for fatigue crack propagation in grey cast iron. As a result, fatigue crack propagation curves for various samples of cast iron including cast iron used in WDNs have been obtained. Figure 4.4 presents a summary of representative grey cast iron fatigue crack propagation curves obtained from published studies (Mohebbi et al., 2010; Speidel, 1981; Bulloch, 1995; Baicchi et al., 2007; James

& Wenfong, 1999). These studies show large variations between laboratory results for fatigue crack propagation in cast iron pipes. The axes in Figure 4.4 are in a logarithmic scale, which further amplifies the large variations in the fatigue curves.

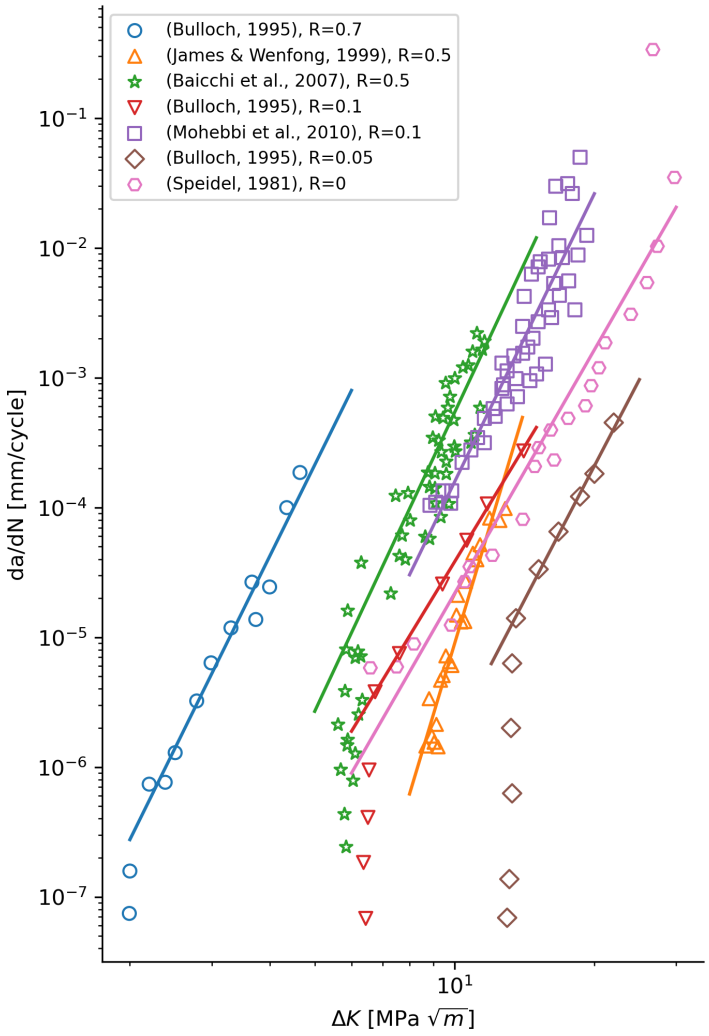


Figure 4.4: Fatigue crack propagation curves for cast iron at different loading ratios R and sources. Loading ratio R is the quotient between minimum and maximum loading in a loading cycle. K stands for stress intensity factor (see Equation 4.1). *Ref stands for reference (Bulloch, 1995; James & Wenfong, 1999; Baicchi et al., 2007; Mohebbi et al., 2010; Speidel, 1981). After: Jara-Arriagada & Stoianov (2024).

A key difference between these curves is their spread locations within the x-axis. The main cause for these deviations in the curves, apart from the heterogeneity that can exist between the cast iron samples, is the loading ratio R at which the fatigue experiments were conducted. The loading ratio R is the quotient of the minimum (valley) loading divided by the maximum (peak) loading of a loading cycle, and serves as a surrogate of

the mean pressure. For a same magnitude of a cyclic pressure (stress) range, a higher loading ratio R defines that the cyclic pressure range is applied under a higher mean pressure.

As observed in Figure 4.4, for a same referenced material, higher R ratios (e.g. higher mean pressures) under a constant cycle range lead to a faster fatigue crack propagation. An explanation for this phenomenon is the theory for crack closure proposed by Elber (1970). The basics of this theory is that unless a crack is affected by a minimum loading called K_{op}, the crack will stay closed, and pressure fluctuations will not cause fatigue. This means that under a low mean pressure, only a portion of the full range of a pressure cycle (ΔK_{eff}) contributes to fatigue crack propagation. In comparison, under higher mean pressures, the full range of a cyclic loading (ΔK) is contributing to fatigue crack propagation. Various modelling theories exist to account for the dependence of a fatigue crack propagation and the loading ratio R. Section 4.3.5 describe some of these, and the model used in this study.

Published studies such as the one in Mohebbi et al. (2010) have applied fracture mechanics to assess the fatigue life of pipes. In contrast to the current study, the authors focused on the role of pressure-induced fatigue in the whole life span of pipes. Under such consideration, it is expected that the contribution of fatigue in crack growth may be rather limited given the strong dependence of fatigue to the initial crack size. However, when considering highly deteriorated pipes, as most cast iron pipes currently in operation are, fatigue could be a major damaging factor that can accelerate failure. In addition, most published analyses, have performed fatigue estimations using fatigue crack propagation curves obtained at R ratios of 0.1. Such R ratio does not represent actual R ratios observed for water pipes in operation. In reality, it has been observed that R ratios in WDNs can be as high as 0.7. This high R ratio indicates that loading conditions do not necessarily need to be close to the critical load to generate fatigue on pipes. Therefore, fatigue life estimations should consider more realistic scenarios for fatigue crack propagation curves.

Furthermore, corrosion can also shift the fatigue crack propagation curves for cast iron

pipes. Corrosion can produce a synergistic combination with fatigue causing accelerated fatigue in pipes at lower pressure-induced stress. Studies such as the one of Cullin et al. (2015) have shown evidence of this effect. However, no experimental studies that include the acceleration of fatigue crack growth due to corrosion have been performed. A common practice to model corrosion fatigue is to treat the rate of propagation of cracks due to fatigue and the rate of propagation of pits due to corrosion as separate processes that add together (Larrosa et al., 2018). However, this does not capture the synergistic effect of both processes. The effect of corrosion indicates that results from pure fatigue crack propagation analysis must be considered as a suggestion of the minimum damage that a cast iron pipe could be exposed to.

4.2.6 Observed failures in cast iron pipes

Observed failures in cast iron pipes are well documented and presented in published research papers (Makar, 2000; Rajani & Kleiner, 2001; Rajani, Lewandowski & Margevicius, 2012). Common failure types include circumferential failure, bell split, longitudinal failures, corrosion holes and pipe blowouts (Makar, 1999). It has been observed that circumferential failures are more frequent in water distribution pipes, while longitudinal failures are more frequent in large diameter transmission mains (Makar, 1999). These observations, however, are highly dependant on the analysed WDN. Also, the described failure modes commonly refer to the shape of the observed fracture after inspection of a failed pipe (e.g. Figure 4.8), and do not describe the mechanism at which a pipe actually failed (Makar et al., 2001).

Various examples of multistage failures in cast iron have also been reported in the literature (Makar, 2000; Rajani, Lewandowski & Margevicius, 2012). Makar (2000) presented a post-mortem analysis of an exhumed pipe. The analysis concluded that a corrosion hole initiated a crack, which gradually extended until the pipe failed. Within the same study another analysis on a circumferential failure showed a similar pattern; namely, a corrosion hole initiated a crack, which then rapidly propagated. These two-stage failures

had different chemical compositions within the hole and the crack surfaces. The corrosion hole exhibited the presence of elements from the surrounding soil, which indicates long-standing existence. In comparison, the crack surface did not show the presence of these elements, which suggests that the crack had developed in a shorter time period before the failure (Makar, 1999). Furthermore, three-stage failures were observed in some of the analysed samples of cast iron pipes (Makar, 1999).

In addition to the published studies presented in this section, over 3000 pipe failure reports were reviewed as part of this investigation. These reports were provided by a water utility in England. The reports included photographs of exhumed failed cast iron pipes taken from a period of approximately 10 years from 1996 to 2006. Most reports included a view of the pipe cross-sectional area before sandblasting and the longitudinal section after sandblasting (Figure 4.5). Various pipes showed a large degree of deterioration caused by corrosion as corrosion pits and patches (Figure 4.6). Most recurrent observed damage were corrosion pits. Some pipes that presented corrosion holes also showed longitudinally propagated cracks (Figure 4.7). Various pipes also showed signs of a failure caused by the blowout of a corrosion patch, possibly because of the propagation of a longitudinal crack within the patch (Figure 4.8).

In Figure 4.6, various sections of the presented pipes are highly reduced by corrosion, however, without the presence of cracks, the pipes only show small localised corrosion holes, without major blowouts. In contrast, it seems that in Figure 4.9, the presence of a longitudinal crack within the deteriorated corroded section of the pipe aided in the final blowout of the corroded section. Moreover, in most cases of pipe blowouts, the largest dimension of the blowout is in the longitudinal section (Figure 4.8). This evidence seem to support the hypothesis that the propagation of a crack within a corrosion patch cause the breakage of the pipes. This scenario is considered in this investigation.



(a)



(b)

Figure 4.5: Example of an exhumed pipe provided by the water utility. a) Cross-sectional section before sandblasting. b) longitudinal section view after sandblasting. Samples provided by Bristol Water, UK. From: Jara-Arriagada & Stoianov (2024).



Figure 4.6: Examples of pipe deterioration due to corrosion pits and patches after sand-blasting pipes. Samples provided by Bristol Water, UK. From: Jara-Arriagada & Stoianov (2024).



Figure 4.7: Longitudinal failures in water distribution pipes due to initial corrosion pits. Samples provided by Bristol Water, UK. From: Jara-Arriagada & Stoianov (2024).



Figure 4.8: Examples of cast iron pipe blowouts. Samples provided by Bristol Water, UK. From: Jara-Arriagada & Stoianov (2024).



Figure 4.9: Pipe blowout with the presence of longitudinal crack. Samples provided by Bristol Water, UK. From: Jara-Arriagada & Stoianov (2024).

4.3 Assessment of water pressure fluctuations on fatigue failures in cast iron pipes using fracture mechanics

This chapter theoretically assesses the impact of water pressure fluctuations on fatigue failures in ageing cast iron pipes. This assessment allows to hypothesise whether and under which conditions a reduction in water pressure fluctuations in a WDN can reduce failure rates and extend the life of cast iron pipes at the onset of and during their wear-out phase. The potential for fatigue failure was evaluated with the use of fracture mechanics fatigue crack propagation in combination with data gathered for material properties, the reduction in pipe wall thickness, and pressure loading conditions observed in water distribution networks. The fracture mechanics analyses assume that cracks already exist in pipes from degradation processes such as corrosion and long term use. This section describes the modelling theories used for the applied mechanical analyses.

4.3.1 Loading conditions for water distribution pipes

The main loading conditions for buried water distribution pipes include internal water pressure, external bending forces, external forces that can crush a pipe, and forces caused by ground movement and thermal expansion (Makar et al., 2001). It has been estimated that the normal operational internal water pressure can be between 6 to 7 times greater than common external crushing loadings (Robert et al., 2022).

The magnitude and dynamic behaviour of water pressure in water distribution pipes continuously vary because of diurnal changes in water demand as well as the operation of pump and valve controls. The stochastic water demand and variations in controls cause unsteady-state pressure fluctuations. Studies of the dynamic pressure fluctuations in operational water distribution networks have shown that pressure fluctuations might either be rapid events also known as pressure transients and/or gradually developing

events (Hoskins, 2015; Rezaei, 2017). Further monitoring campaigns in operational water distribution networks have also shown that, although not ubiquitous, frequent pressure fluctuations over 30 mH_2O may commonly occur (UKWIR, 2022). This magnitude of pressure variations pose a risk for fatigue induced breaks in pipes. Figure 4.10 shows an example of an observed pressure transient event in a water distribution pipe logged by an UK water utility. Based on the gathered pressure data, the impact of constant amplitude loadings will be analysed at 20 mH_2O, 30 mH_2O, 40 mH_2O and 50 mH_2O water pressure head on fatigue failures in ageing cast iron pipes.

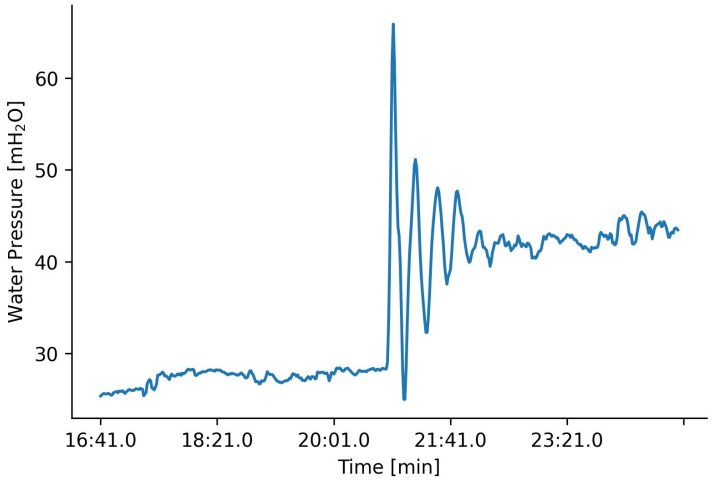


Figure 4.10: Pressure transient in a water distribution pipe acquired from an operational network. From: Jara-Arriagada & Stoianov (2024).

4.3.2 Distributions of size and pipe wall thickness for cast iron pipes

The manufacturing of cast iron pipes has been regulated in the UK by British Standards (e.g. BS 78, BS 1211, BS 2035, BS 4622), which had multiple revisions over the years. For this investigation, pipe sizes and wall thicknesses provided by British Standard 78:1917 were used for pit cast iron pipes and British Standard 1211:1958 for spun cast iron pipes. Figure 4.3 shows the wall thickness by the diameter size for different classes of cast iron pipes. The classes define the design pressure that the pipe should withstand.

4.3.3 Stress variations due to a reduction in pipe wall thickness

As discussed in section 4.2.4, pitting corrosion is a pervasive issue in cast iron pipes. In some instances, corrosion pits amalgamate and create large corrosion patches that reduce significant portions of a cast iron pipe section, and increase the stresses in the pipe (Soltani Asadi & Melchers, 2017*b*; Rathnayaka et al., 2017). Recently, Fu et al. (2020) proposed an empirical equation based on a finite element analysis to predict the stress concentration factor for pipes with different size corrosion patches. Their analysis showed wide variations in stress concentration factors. For example, they showed that stress concentration factors for corrosion patches can vary in between 3 to 9 for different ranges of the length/width ratio and between 10 to 5 for 2×(width/depth) ratios greater than 5. The analysed factors suggest that the assumption of uniform wall thickness reduction using thin wall cylinder theory as done for the presented analysis is adequate for assessing the impact of pressure fluctuations on pipes with reduced wall thickness (see Figure 4.13). Therefore, in order to assess the impact of pipe wall thickness reduction on ageing cast iron pipes, the stresses in a pipe are estimated using thin wall cylinder theory for different pipe wall thicknesses.

4.3.4 Fracture mechanics analysis

As discussed in sections 4.2.2 and 4.2.3, the microstructure of cast iron pipes would lead to the development of cracks. Even a new cast iron pipe can have graphite flakes up to 2 mm, which can produce a similar effect to a crack (Baker, 1978). Corrosion pits can also simulate sharp cracks in some cases. Under these circumstances, the use of fracture mechanics is the most suitable approach to assess the resistance of a pipe to external loadings. This section presents an overview of the fracture mechanics analysis performed in this chapter.

In fracture mechanics, single or multiple parameters are used to describe the stress intensification in a material due to the presence of cracks and loadings (Anderson, 2017). The

single or multiple parameters can be compared to the maximum value that a material can withstand, which can be determined experimentally, and be generalised for the material as long as its properties remain unchanged. The type of single or multiple parameters to be used in fracture mechanics depend on the level of yielding that a load generates near the tip of the crack. In cases of very low or no yielding, the concept of stress intensity factor can be applied (Irwin, 1956). This research makes use of the stress intensity factor.

The stress intensity factor K [MPa$\sqrt{m}$] can be defined as follows:

$$K = Y \cdot \sigma \cdot \sqrt{\pi \cdot a} \tag{4.1}$$

where Y is a dimensionless geometry function that depends on the shape of a material and the crack length, σ is the stress applied to the material, and a is the crack length or half the crack length depending on the formulation. The value of the stress intensity factor represents a comparable numerical magnitude of the stress field near the tip of a crack. The stress intensity factor can be compared to the fracture toughness of the material, which is a measure of the maximum stress intensity factor that a material can withstand (Anderson, 2017).

In case of complex geometries, the value of the stress intensity factor can be also estimated by a finite element analysis. For pipes, general empirical expressions have been proposed and are commonly applied. Such expressions include longitudinal through cracks, longitudinal surface cracks, and penny-shape cracks. Some research studies have evaluated more detailed formulations by simulating stress intensity factors in a pipe due to sharp cracks (Wang et al., 2017). These studies have shown that the results of the prior empirical formulation can overestimate results from more detailed analyses. Nevertheless, such an overestimation can be used as a factor of safety in the calculations. Given the various formulations, this investigation utilises fracture mechanics calculations based on the formulation of a longitudinal through crack in a cylinder (Newman, 1976).

In particular, the stress intensity factor formulation was used to determine the maximum

crack length for a break in a pipe. This stress intensity factor formulation was also used to determine values for minimum water pressure fluctuations required to start the process of fatigue, and in the fatigue crack propagation analysis. Maximum crack lengths were calculated by solving the value of crack length that produces a stress intensity factor which reaches the fracture toughness of a cast iron pipe for different values of water pressure. A similar approach was performed to obtain the minimum magnitude of cyclic water pressure required to start causing fatigue crack propagation in a cast iron pipe. In this case, the value of water pressure that generates a stress intensity factor equal to the stress intensity threshold for different values of crack lengths was calculated. The procedure for the fatigue crack propagation analysis to estimate fatigue-induced failure is presented in the next section.

4.3.5 Fatigue crack propagation

Fatigue crack propagation deals with the propagation of cracks due to cyclic loadings. In particular, the fatigue crack propagation theory is based on the observation that, under controlled laboratory conditions, the rate of crack growth per loading cycle (da/dN) is correlated with the magnitude of the stress intensity range (ΔK) exerted on a material per cycle (Figure 4.11). This stress intensity range is the range of the cyclic loading applied converted to stress intensity factor.

Important information derived from Figure 4.11 includes: (1) The curve for fatigue crack growth can be expressed in general terms as presented in Equation 4.2. (2) For every material, there is a minimum stress intensity threshold range (ΔK_{th}), below which, there is no fatigue crack propagation. (3) Three regions are observed in a fatigue crack propagation curve. Region two is the common region used for fatigue crack propagation analysis, as loading conditions usually lie in this region. (4) The steepness of the curve slope, m, provides indication of the usefulness of fatigue crack propagation assessment for different materials. If the slope is too steep, then there is little difference between the minimum load to start causing fatigue (K_{th}) and the maximum load to break a material (K_c). In

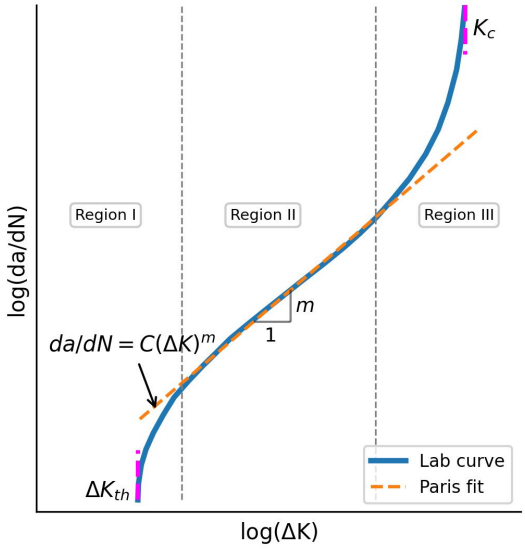


Figure 4.11: Illustration of the empirical relation between the rate of crack propagation (da/dN) and stress intensity factor range (ΔK). From: Jara-Arriagada & Stoianov (2024).

this chapter, it is considered that the slope of the fatigue crack propagation curves for cast iron pipes presented in Figure 4.12b, at approximately 6.8, provides sufficient difference between K_{th} and K_c for fatigue crack propagation analysis.

$$\frac{da}{dN} = f\left(\Delta K, R, \ldots\right) \tag{4.2}$$

On the basis of the empirical relation shown in Figure 4.11, the proposed functional form of Equation 4.2, and also assuming a constant cyclic loading, the number of cycles, N, required for a crack to reach a specific final crack can be estimated as follows:

$$N = \int_{a_i}^{a_f} \frac{1}{f\left(\Delta K, R, \ldots\right)} da = \int_{a_i}^{a_f} \left(\frac{dN}{da}\right) da \tag{4.3}$$

Fatigue crack propagation laws

To utilise the functional form of Equation 4.2, several detailed expressions for the function of da/dN have been proposed. A basic expression is the Paris law (Equation 4.4). This

equation relates the rate of crack propagation with the stress intensity range in the region II of the crack propagation curve in Figure 4.11:

$$\frac{da}{dN} = C \cdot \Delta K^m \tag{4.4}$$

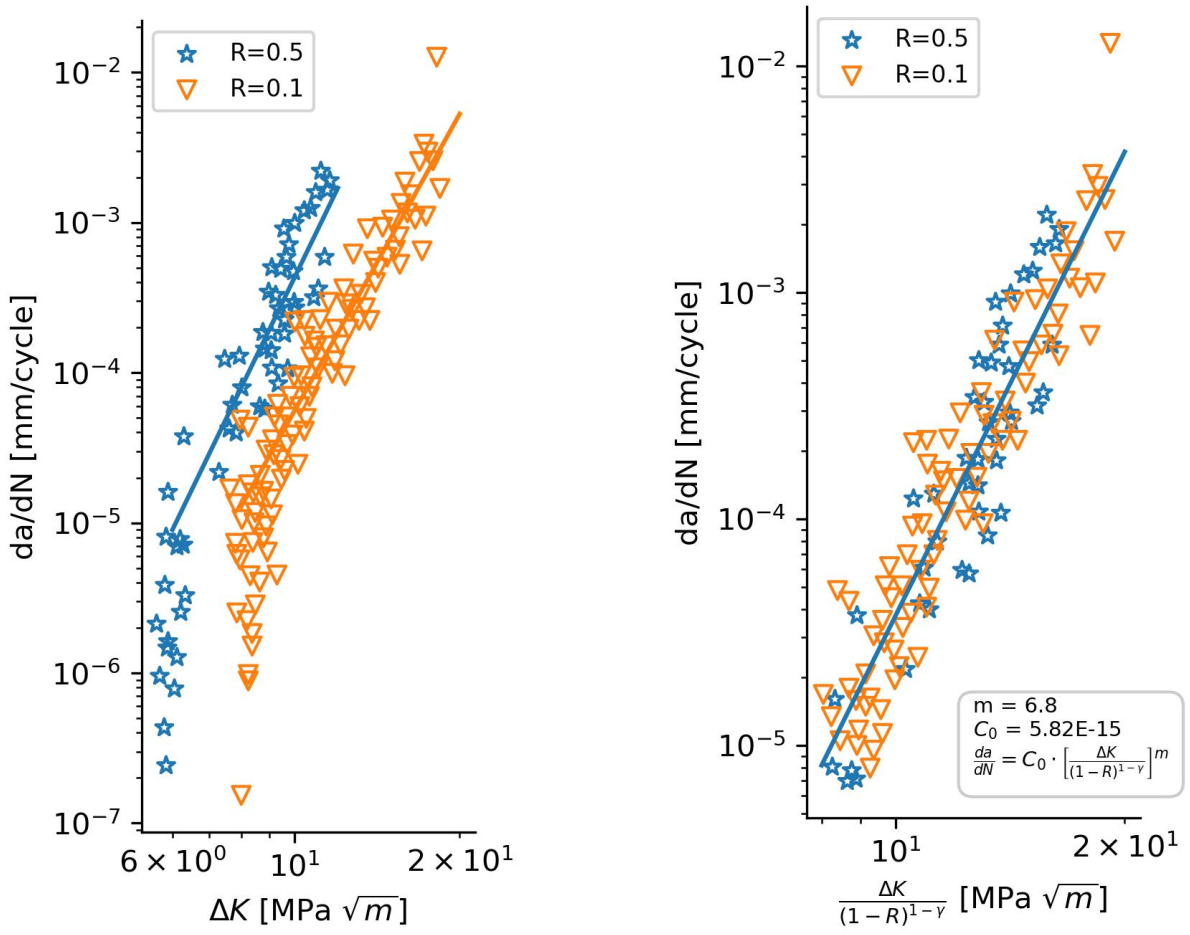
The Paris law is commonly used since it describes well the region II in Figure 4.11. However, this equation only works on a fix loading ratio R, which is the loading ratio used to obtain the fatigue crack propagation curve. A more general expression to assess fatigue crack propagation for different loading ratios Rs is the Walker equation (Equation 4.5) (Walker, 1970).

$$\frac{da}{dN} = C_0 \cdot \left[\frac{\Delta K}{(1-R)^{1-\gamma}}\right]^m \tag{4.5}$$

Similarly to the Paris law, the Walker equation only describes fatigue crack propagation in region II. The most complete mathematical fit for the fatigue crack propagation curve presented in Equation 4.2 and Figure 4.11 is the NASGRO equation (Equation 4.6). This equation provides a result for any stage of the fatigue curve, but requires multiple constants to be fitted.

$$\frac{da}{dN} = \frac{C\,(1-f)^n\,\Delta K^n\left(1-\frac{\Delta K_{th}}{\Delta K}\right)^p}{(1-R)^n\left(1-\frac{\Delta K}{(1-R)K_c}\right)^q} \tag{4.6}$$

Although additional factors such as temperature can play a role in a fatigue crack propagation analysis, the conducted analysis does not intend to provide exact estimations for fatigue life, which would be extremely difficult. The main aim of this chapter is to assess the general magnitude and the effects of varying water pressure fluctuations and the mean pressure on deteriorated cast iron pipes. Therefore, it was considered that the use of the Walker equation fitted for a selected representative cast iron material was adequate to estimate the rate of crack propagation in ageing cast iron pipes.



(a) (b)

Figure 4.12: Walker equation fit for a selected representative cast iron material. a) Selected fatigue crack propagation data from Baicchi et al. (2007). b) Fitted Walker equation to data. From: Jara-Arriagada & Stoianov (2024).

Fitting the fatigue crack propagation equations

From the fatigue crack propagation curves presented in section 4.2.5, the data provided in reference Baicchi et al. (2007) was selected to fit the Walker equation. This data included fatigue crack propagation curves for cast iron at loading ratios of 0.5 and 0.1 (Figure 4.12a). This data was fitted for the Walker equation with a $\gamma = 0.5$ to obtain a fatigue crack propagation curve that can be used for any loading rate scenario (Figure 4.12b), and therefore to perform an analysis for varied cases of pressure fluctuations and mean pressure.

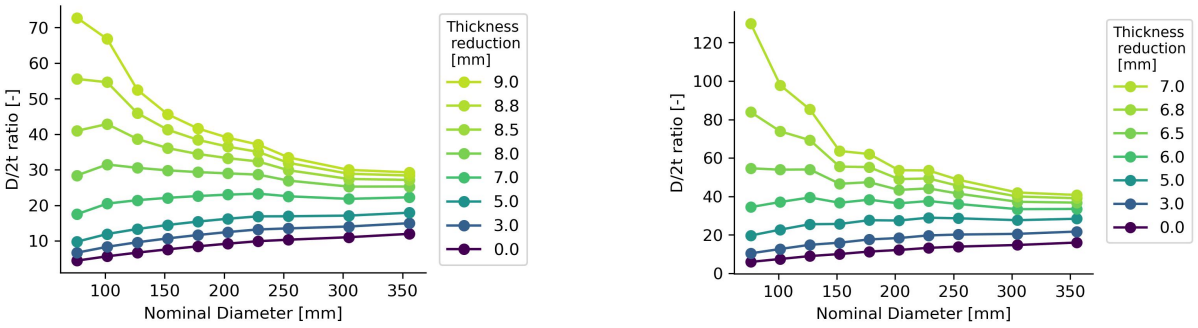
4.4 Results from fracture mechanics evaluation and fatigue crack propagation for cast iron pipes

This section presents the results of the theoretical calculations, which were carried out to investigate the impact of water pressure fluctuations on fatigue failures in deteriorated cast iron pipes. An analysis of the effect of the reduction in pipe wall thickness on the hoop stresses in cast iron pipes was first conducted to assess the detrimental effect of corrosion patches. Then, fracture mechanics analysis were carried out to understand the limits for maximum crack lengths, and minimum magnitudes of water pressure for cast iron pipes to start experiencing fatigue failures. These results inform the fatigue crack propagation analyses, which show the conditions under which a reduction in water pressure fluctuations can considerably affect the fatigue life of a pipe. The analyses were performed for both pit and spun cast iron pipes with the exception of the fatigue crack propagation analysis, which was done for pit cast iron pipes because of the available data.

4.4.1 Effect of the reduction in pipe wall thickness on pipe stresses

This analysis investigates the increase in hoop stresses with the reduction in pipe wall thickness. The values for the considered pipe wall thicknesses were based on the British Standards 78:1917 and 1211:1945 for pit and spun cast iron pipes. The plots in Figure 4.13 show how the ratio $D/2t$ changes for different diameters, and how wall thickness reduction affect its magnitude. According to the thin wall cylinder theory, hoop stresses can be approximated in a pipe by multiplying the stress produced due to water pressure by the factor $D/2t$. Therefore, the plots in Figure 4.13 show how the magnitude of applied stresses vary due to water pressure fluctuations for different pipe diameters and values for the reduction in pipe wall thickness.

The results show that there is a considerable increase in applied stresses especially for



(a) (b)

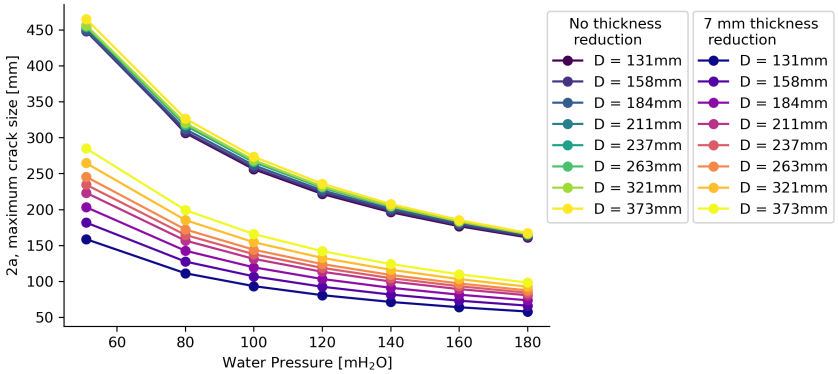
Figure 4.13: Effect of pipe wall thickness reduction on the D/2t ratio for different cast iron pipe diameters. a) Pit cast iron pipe. b) Spun cast iron pipe. From: Jara-Arriagada & Stoianov (2024).

small diameter pipes following a reduction of 6 mm and 8 mm in pipe wall thickness in spun and pit cast iron pipes respectively. This rate of increase in stresses, however, is less pronounced for pipes with diameter above 250 mm, such as transmission mains. This basic analysis indicates that stresses can easily increase between 4 and 6 times considering the typical corrosion patches observed in ageing pipes as presented in section 4.2.4.

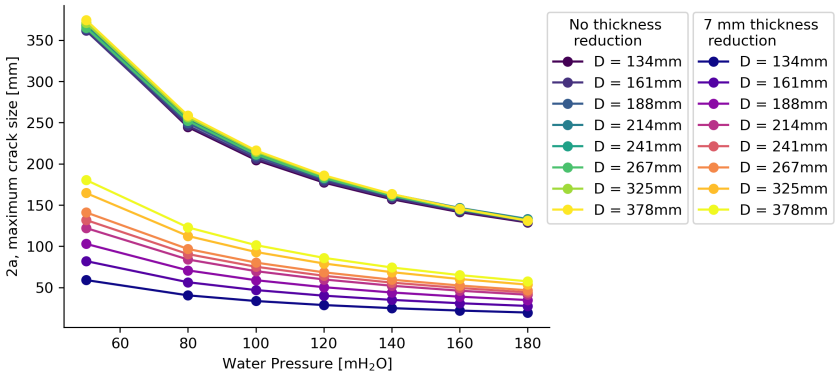
4.4.2 Maximum crack lengths in cast iron pipes

This analysis investigates the maximum theoretical crack length at which a crack can propagate before the fracture of a material. This maximum crack length affects the number of pressure cycles required to break a pipe from an initial crack length. Therefore, the values presented here inform the calculation for the fatigue crack propagation estimations further presented in section 4.4.4.

The plots in Figure 4.14 show the maximum crack length before the break of a pipe when different magnitudes of water pressure are applied. This crack length was calculated by solving the value of crack length which produces a stress intensity factor (Equation 4.1) equal to the fracture toughness of the pipe for a given magnitude of water pressure. Values of 18.1 MPa$\sqrt{\text{m}}$ and 20.36 MPa$\sqrt{\text{m}}$ were assumed for the fracture toughness of pit and spun cast iron pipes respectively, on the basis of the references provided in Table 4.1. The analysis is done for different pipe diameters, which are taken from British Standards



(a)



(b)

Figure 4.14: Maximum crack length for cast iron pipes as a function of increasing water pressure and pipe wall thickness reduction. a) Pit cast iron pipe. b) Spun cast iron pipe. After: Jara-Arriagada & Stoianov (2024).

(section 4.3.2). The applied water pressure ranges from 50 mH_2O to 180 mH_2O. Two clusters of plotted lines can be observed in Figure 4.14a and Figure 4.14b. The clusters of plotted lines at the bottom show the maximum crack lengths when considering a reduction in pipe wall thickness of 7 mm for the different pipe diameters. The upper clusters of plotted lines show the expected maximum crack lengths for pipes with intact pipe walls. Note that the maximum crack length calculated for a pressure of 50 mH_2O for a pipe with a reduced wall thickness agrees with the approximate maximum crack length observed in a physical fatigue experiment performed by Wong et al. (2017).

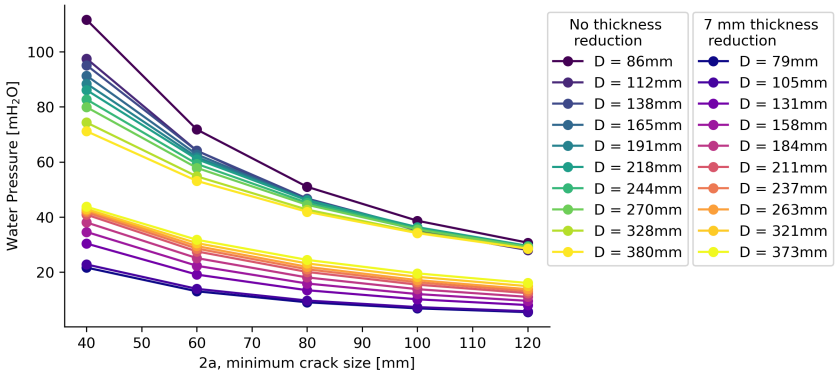
The results presented here show the trade-off between pipe wall thickness and material resistance for spun and pit cast iron pipes. Although spun cast iron pipes have a greater tensile strength and fracture toughness, the reduced wall thickness of spun cast iron pipes makes these pipes more prone to failures in comparison with pit cast iron pipes. In

addition, the presented results can be used to indicate a maximum crack size that could withstand a maximum expected pressure transient in a network without resulting in a pipe break. For the fatigue analysis presented in this paper, it has been assumed that a pipe should withstand a maximum pressure transient of 100 mH_2O. Although not frequent, such magnitude of a pressure transient occurs in water distribution pipes as reported by the UKWIR (UKWIR, 2022).

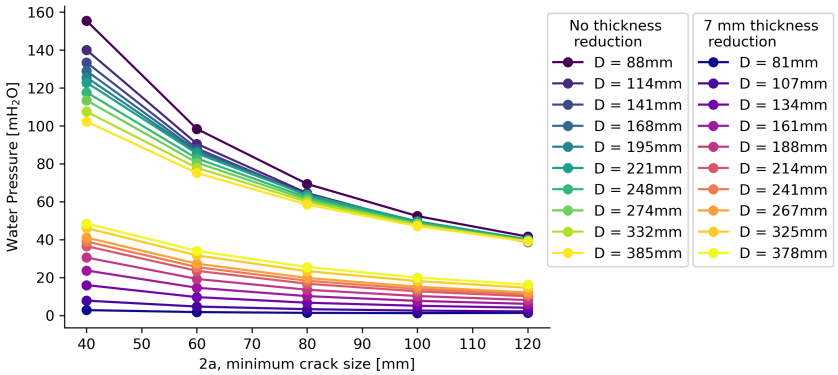
4.4.3 Minimum magnitude of water pressure cycles for pressure-induced fatigue

This section describes the results of the minimum magnitude of water pressure cycles required to initiate fatigue crack propagation in cast iron pipes. The magnitude of the pressure cycles are combined with a crack length that produces a stress intensity factor equal to the stress intensity threshold of a pipe (ΔK_{th}). The value of ΔK_{th} highly depends on the load ratio R used in the fatigue experiment to determine the fatigue crack propagation curve of a material. Most fatigue crack propagation curves for cast iron have been determined using a loading ratio R between 0.1 and 0.3, although the load ratio R in WDNs can exceed 0.5. The difference in loading ratios Rs can produce substantial differences in calculations for the stress intensity factor for fatigue and therefore in the life expectancy of a pipe exposed to fatigue cycles.

In order to compare results for both pit and spun cast iron pipes, a minimum ΔK_{th} of 2 MPa$\sqrt{m}$ (at loading ratio R of 0.7 Figure 4.4) was used for pit cast iron and 4 MPa$\sqrt{m}$ for spun cast iron (assumed). Figure 4.15 shows that when pipes are not affected by a reduction in wall thickness, higher magnitudes of water pressure cycles are required in spun cast iron pipes than pit cast iron pipes to exceed the stress intensity threshold over which fatigue crack propagation occurs. However, as the wall thickness is reduced, the spun cast iron pipes become more susceptible to fatigue failures than pit cast iron pipes. A key observation from both plots in Figure 4.15 is that for pipes not affected by a reduction in wall thickness, commonly observed pressure cycles in water distribution pipes (∼10-



(a)



(b)

Figure 4.15: Minimum water pressure to reach stress intensity threshold for cast iron pipes. The effect of reduction of thickness is estimated. a) Pit cast iron pipe. b) Spun cast iron pipe. After: Jara-Arriagada & Stoianov (2024).

30 mH_2O pressure head) are unlikely to cause fatigue failures even for scenarios when a long crack had been generated by an external factor. Only pipes with long cracks and under large pressure fluctuations could be affected by pressure-induced fatigue if there is no reduction in pipe wall thickness. However, most pipes at the onset of and during their wear-out phase are commonly affected from a high level of deterioration, which significantly reduces the minimum water pressure for pressure-induced fatigue.

4.4.4 Fatigue crack propagation

This section presents the results from fatigue crack propagation analyses on a representative ~150 mm pit cast iron pipe. The main goal is to assess conditions under which pressure-induced fatigue can cause failures in deteriorated cast iron pipes, and how to

reduce their rate of fatigue deterioration.

Figure 4.16 shows the effect of pipe wall thickness on the number of cycles to failure for different initial crack lengths and pressure fluctuations at a constant loading ratio R of 0.5. Two cases illustrated by the areas in colour were defined. The upper scenario considers pipes with no reduction in wall thickness. The scenario at the bottom of the plot considers a reduction in pipe wall thickness of 7 mm caused by a corrosion patch. The x-axis indicates an increasing crack length, and the y-axis represents the number of cycles of water pressure at constant magnitude to reach the maximum crack length before a break in a pipe. For every coloured area, the lower bound represents a constant cyclic pressure of 50 mH_2O and the upper bound represents a constant cyclic pressure of 20 mH_2O. The minimum water pressure cycle to cause fatigue was not included in this analysis for cases when the pressure is below a fatigue threshold. This omission was performed for clarity.

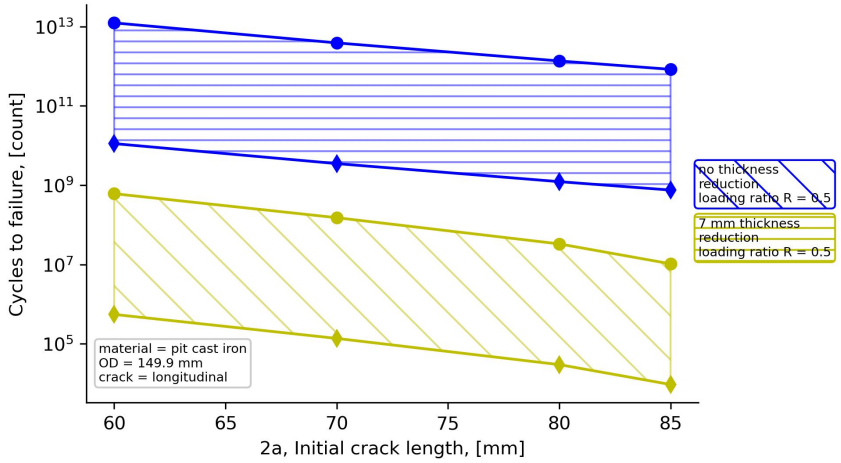


Figure 4.16: Effect of pipe wall thickness on the number of cycles to failure for different initial crack lengths and water pressure fluctuations. From: Jara-Arriagada & Stoianov (2024).

From the two presented scenarios, it can be observed that only the scenario that considers a reduction in pipe wall thickness is of relevance for this investigation. Pipes that are unaffected by a reduction in wall thickness are unlikely to be damaged by fatigue as the required number of cycles largely exceeds observed pressure cycles in WDNs. For example, cyclic pressure loadings in the range of 20 mH_2O to 50 mH_2O are observed

to normally vary between 2000 to 8000 cycles per year approximately, as presented in the next section. Given this number of observed pressure cycles and magnitudes, it is concluded that only the scenario, which considers a reduction in pipe wall thickness can lead to pressure-induced fatigue.

Figure 4.17 illustrates the scenario with a reduction in pipe wall thickness, and how changes in pressure cycle range and mean pressure affect the number of cycles to break a cast iron pipe. The lower curve shows the number of cycles needed to break a pipe under constant cycle ranges of 50 mH_2O and a mean pressure of 60 mH_2O. The successive upper lines show how a combined reduction of both pressure cycle range and mean pressure affects the number of cycles to break a pipe. It can be observed that the initial crack length plays a critical role in the number of cycles to failure, and pressure-induced fatigue is relevant for pipes that are about to break from a few 9 mm to 4 mm of crack increase. Figure 4.17 also show that the reduction of both pressure cycle range and mean pressure can significantly increase the number of cycles to break a pipe.

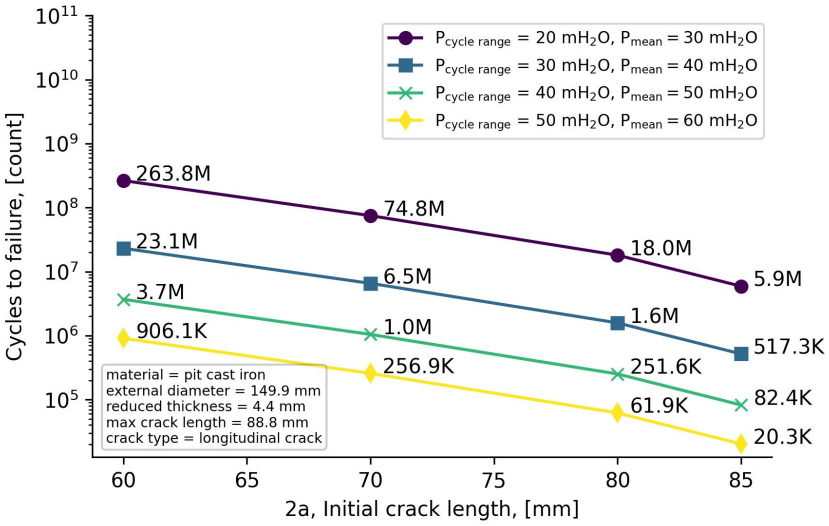


Figure 4.17: Fatigue life estimation with a reduction in pipe wall thickness at varied cyclic pressure loadings and mean pressure. From: Jara-Arriagada & Stoianov (2024).

Figure 4.18 and Figure 4.19 show the individual contributions of reducing the magnitude of pressure fluctuations and mean pressure on the fatigue life of a pipe. Figure 4.18 shows various scenarios for the number of cycles needed to break a pipe by varying the mag-

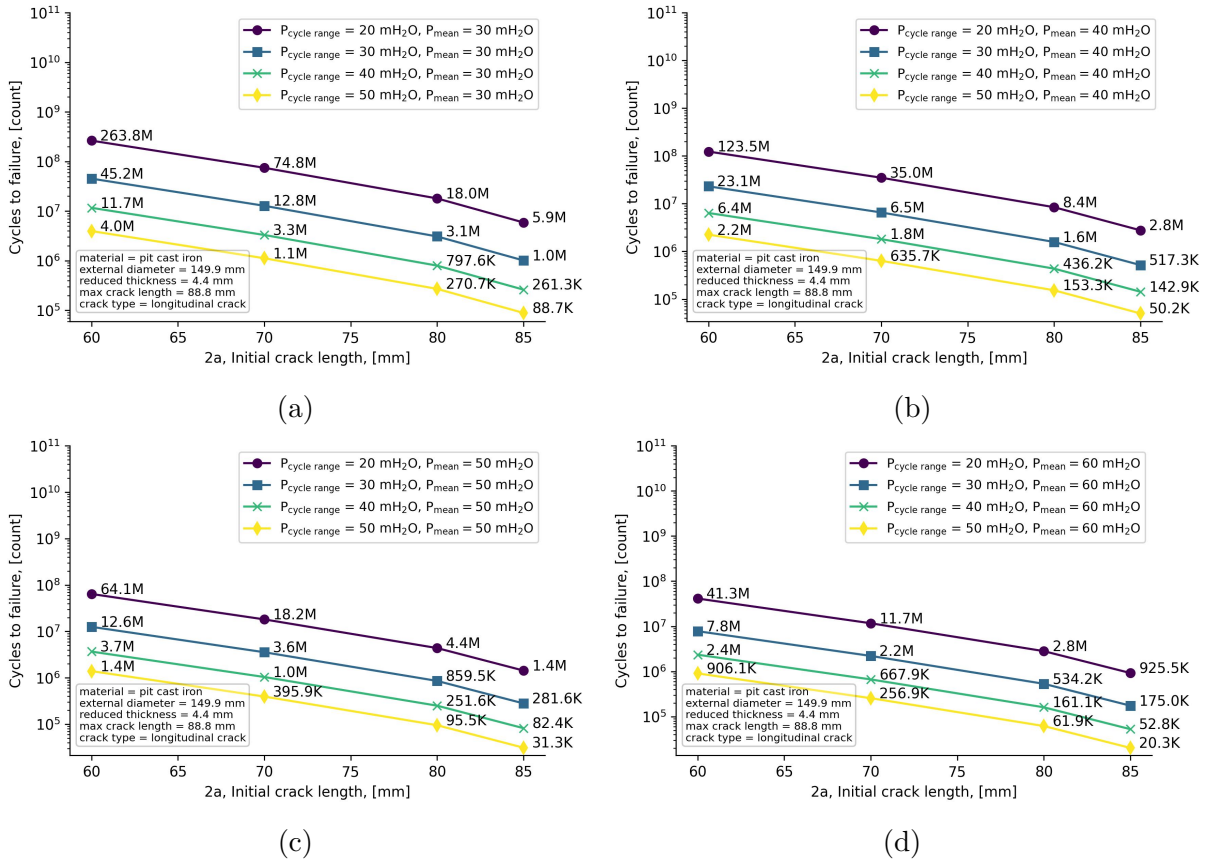
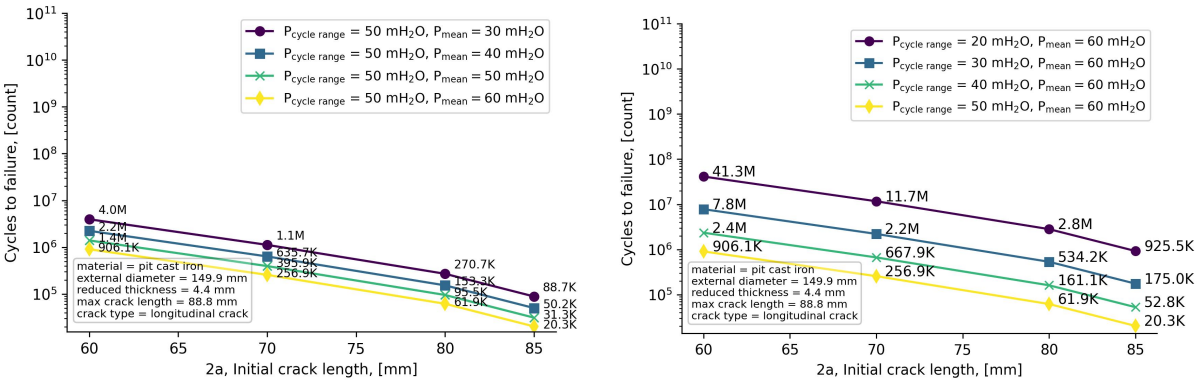


Figure 4.18: Fatigue life estimations for varying magnitudes of pressure fluctuations at constant magnitudes of mean pressure. a) Mean pressure at 30 m. b) Mean pressure at 40 m. c) Mean pressure at 50 m. d) Mean pressure at 60 m. From: Jara-Arriagada & Stoianov (2024).

nitude of water pressure fluctuations under a constant magnitude of the mean pressure. Figure 4.19 compares the individual contribution of only reducing pressure fluctuations versus only reducing mean pressure from a starting case of 50 mH_2O of pressure cycle range and 60 mH_2O of mean pressure. From this figure, it can be observed that reduction of the magnitude of pressure fluctuations is more impactful than reduction of mean pressure to increase the life of deteriorated pipes. For the proposed case, a reduction of mean pressure from 60 mH_2O to 30 mH_2O under a constant pressure cycle magnitude increases in excess of 4 times (4.37) the number of cycles required to break a pipe. In contrast a reduction of the magnitude of cycles from 60 mH_2O to 30 mH_2O increases the number of cycles to failure in almost 9 times (8.62).

Nevertheless, when the individual effects of reducing both the magnitude of pressure fluctuations and mean pressure are compared to the effect of reducing both pressure



(a) (b)

Figure 4.19: Comparison cycle range reduction versus mean pressure reduction. a) Varied mean pressure and constant magnitude of cycle range. b) Varied magnitudes of cycle range and constant mean pressure. From: Jara-Arriagada & Stoianov (2024).

metrics at the same time, the difference is considerable. From Figure 4.17, it is observed that a combined reduction in the magnitude of pressure cycles from 50 mH_2O to 30 mH_2O, and mean pressure from 60 mH_2O to 40 mH_2O can increase the fatigue life 25.5 times with respect of the original scenario of 50 mH_2O of pressure cycle range and 60 mH_2O of mean pressure. It is also noticeable, that even a small combined reduction of 10 mH_2O of mean pressure and 10 mH_2O of pressure fluctuations has the theoretical potential to quadruple the required number of cycles to failure of the pipe. These examples show that the combination of reducing the magnitude of pressure cycles and mean pressure have a synergistic effect.

Finally, Figure 4.20 presents the what if case of reducing mean pressure but slightly increasing pressure fluctuations. The plot shows that the benefit of mean pressure reduction is outweighed by the negative effect of increasing the magnitude of pressure cycles, leading to less cycles to break a pipe. This finding is important as various schemes for reduction of average zone pressure rely on the use of pressure control devices, such as valves, which, if improperly manipulated, can increase the magnitude and frequency of pressure fluctuations, which could lead to a decreased fatigue life in pipes, and consequently more pipe breaks.

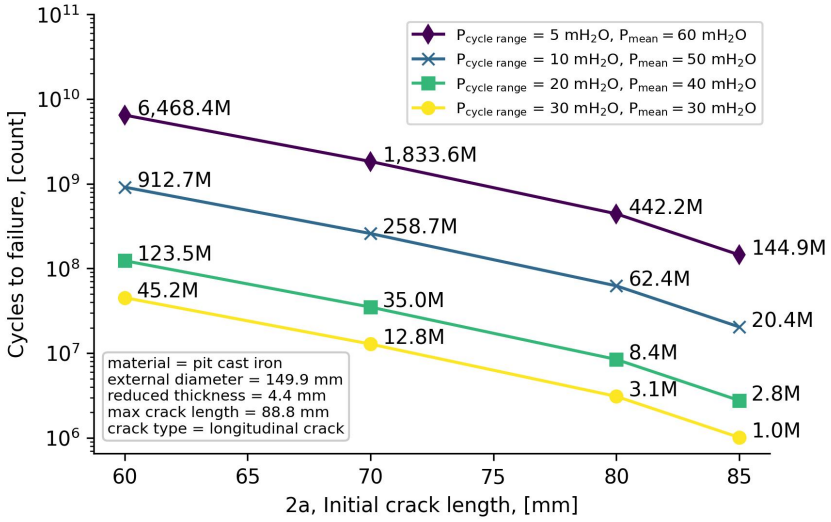


Figure 4.20: Effect of mean pressure reduction and increase of pressure cycles. From: Jara-Arriagada & Stoianov (2024).

4.5 Discussion

This chapter investigated the potential impact of pressure-induced fatigue on breaks of ageing cast iron pipes, which are at the onset of and within their wear-out phase. The research has focused on: 1). examine typical deterioration conditions in cast iron pipes; 2). evaluate the magnitude and frequency of pressure fluctuations that may lead to pressure-induced fatigue failure; 3). quantitatively estimate of the influence of mean pressure on pressure-induced fatigue; and 4). assess the application of advanced pressure control for reducing deterioration in cast iron pipes. These research questions were addressed through a comprehensive review of published literature, the analysis of an extensive and unique collection of pipe break sample reports from a UK water utility, and mechanical analyses using fracture mechanics fatigue crack propagation.

The review for cast iron pipes in WDNs together with the observations from pipe failure reports showed that pitting corrosion is the most recurrent type of deterioration in cast iron pipes. Corrosion pits also usually amalgamate creating large corrosion patches (Figure 4.6). Published research have reported observed depths of pitting corrosion and corrosion patches to be approximately 6 mm to 7 mm in old pipes irrespective of their age. The presence of corrosion patches in deteriorated cast iron pipes was found to be rel-

evant for the occurrence of pressure-induced fatigue. The presented analysis (Figure 4.16) suggests that pipes must have a reduced pipe wall thickness in order for pressure-induced fatigue to significantly contribute to pipe failures. This conclusion complements previous studies, which only considered "new" pipes with intact pipe wall thickness for fatigue calculations in cast iron pipes.

For ageing cast iron pipes with reduced pipe wall thickness, the presented analysis indicates that commonly occurring operating pressure fluctuations can extend existent cracks to a critical length. A key observation from the results presented in Figure 4.17 is that the number of pressure fluctuations required to break a pipe are highly dependent on the initial crack length. Moreover, when a pipe has developed a crack near the critical length from external factors such as corrosion, few millimetres of crack extension could significantly reduce the structural resistance of a pipe. Therefore, operational interventions that reduce the crack growth mechanisms could be highly beneficial to minimise the occurrence of imminent breaks in cast iron pipes.

The influence of the magnitude and frequency of pressure fluctuations, and mean pressure on fatigue induced pipe failures was evaluated and presented in Figure 4.17 to Figure 4.20. These figures provide quantitative insights of the potential changes in fatigue life in a pipe due to variations in the magnitude of pressure fluctuations and mean pressure. The presented estimates were not intended to be precise but to define the trends of the variations in pressure-induced fatigue due to changes in water pressure.

The results suggest that an effective method to reduce pressure-induced fatigue in cast iron pipes is to reduce mean pressure. It was calculated that a reduction in mean pressure of approximately 20 mH_2O (while maintaining constant the magnitude of cycles) can more than double the fatigue life of ageing cast iron pipes. This finding aligns well with the previous chapter showing statistical fits of reducing mean pressure and decreasing pipe breaks. Technologies and optimisation methods for implementing advanced pressure reduction have been discussed and implemented in (Changklom & Stoianov, 2019; Pecci et al., 2019; Ulusoy et al., 2022). An important consideration, however, is that the

reduction of mean pressure must not increase pressure fluctuations. The magnitude of pressure fluctuations is more impactful than mean pressure on pressure-induced fatigue, and a decrease in mean pressure with an increase in pressure fluctuations can lead to reduced fatigue life (Figure 4.20). This consideration might also explain industry experience showing in some occasions increases in pipe breaks even in presence of pressure reduction schemes.

As mentioned, the magnitude of pressure fluctuations has a larger impact than mean pressure on pressure-induced fatigue in ageing cast iron pipes. This is an obvious conclusion since there must exist cyclic loadings due to pressure fluctuations for fatigue to occur. Nonetheless, as shown in Figure 4.19, small reductions in the magnitude of pressure cycles (without reducing the number of cycles) can provide significant reductions on the rate of propagation of cracks in cast iron pipes. The benefits from relatively small reductions in the magnitude of pressure fluctuations show to be more than double of the benefits in reduction of mean pressure for a similar pressure reduction range. However, the control of water pressure fluctuations is a more challenging task. In order to consider a cost-effective solution to reduce such pressure component, it is of critical importance to continuously assess the causes for pressure fluctuations (including pressure transients), and how they propagate and dissipate in WDNs. The continuous monitoring of the hydraulic dynamics in WDNs is therefore critical to have these operational insights.

In the case that it is possible to control water pressure fluctuations, this investigation indicates that a significant benefit can be obtained if both mean pressure and pressure fluctuations are reduced at the same time (Figure 4.17). This means that it is possible to balance the reduction of both pressure components according to operational needs and capabilities of pressure control. The results presented in section 4.4.4 were performed for a representative ∼150 mm pipe with a 7 mm corrosion patch. Practitioners can use the provided methodology in section 4.3 to perform fracture mechanics evaluations for other pipe diameters of interest, and perform similar plots to Figure 4.17 and Figure 4.19 to assess and balance the benefits of pressure control for particular needs and systems.

As indicated, the number of cycles to failure greatly depend on the initial crack length. As a crack increases, the damage caused by a same magnitude of cycle range also increases. According to Figure 4.17, it is possible to suggest that cycles of maximum magnitude of 20 mH_2O at a mean pressure of 30 mH_2O could be tolerated in pit cast iron pipes, as long as these parameters do not change over time. However, given the intrinsic variability of water networks, and the sensitivity of fatigue behaviour, safety factors should be added to this suggestion. This is valuable information to assess and balance pressure management schemes. However, it is important to consider the limitations of the presented study; pressure fluctuations in WDNs are generally not constant, very large cycles can occur in between smaller cycles which can alter the rate of crack propagation; the effect of corrosion fatigue was also not considered. Corrosion fatigue could significantly decrease the fatigue life of a pipe; Besides, the fatigue analysis was performed for one pipe diameter with conservative assumptions, and newer pipe materials have much lower fracture toughness values, which should also be considered in future work. Therefore, it is emphasised that the main contribution of this chapter are the trends in the changes in fatigue life due to variations in the pressure fluctuations and mean pressure, instead of the presented absolute values.

As a separate consideration, the review and analysis of the mechanical differences between pit and spun cast iron have also suggested that spun cast iron pipes are at most risk of failure than older pit cast iron pipes. Although spun cast iron pipes are stronger than pit cast iron pipes, the thinner pipe walls of spun cast iron pipes make these pipes more vulnerable to corrosion patches (Figure 4.6). Such information can serve water utilities to better fine tune selection criteria for pipe replacement and rehabilitation. In addition, count of water pressure fluctuations with their range within a time period, and the associated mean pressure would also greatly benefit pipe break prediction models. Pressure fluctuations can be obtained by the use of high frequency pressure monitoring, and edge data processing on a pressure monitoring device using cycle counting algorithms, such as the rainflow cycle counting.

In summary, this chapter has demonstrated that pressure-induced fatigue can play a crit-

ical role in defining the remaining life of ageing cast iron pipes, which are at the onset of and within their wear-out phase. Such conclusions provide a theoretical support for the concept of cumulative pressure-induced stress. As presented, minimal changes in the length of cracks due to fatigue crack propagation can result in the break of pipes. Due to the large stock of deteriorated cast iron pipes in water distribution networks and current rates of replacement, water utilities should proactively consider and implement cost-effective strategies to reduce pipe breaks. This chapter shows that an appropriate reduction of both mean pressure and water pressure fluctuations could be extremely beneficial to reduce pipe breaks in deteriorating cast iron pipes and allow a phased cost-efficient pipe replacement.

4.6 Conclusion

This chapter has investigated the impact of water pressure fluctuations on pressure-induced fatigue failures in cast iron pipes, which are at the onset and/or within their wear-out phase. An extensive review of the material properties and differences for both pit cast iron pipes and spun cast iron pipes was carried out. Based on this review, and an analysis of reports for failed exhumed cast iron pipes in a water distribution network, various mechanical analyses were performed utilising fracture mechanics. The key conclusions and contributions from this chapter are as follows:

(1) Pressure-induced fatigue failures may impact pipes with reduced wall thickness. In particular pipes observing corrosion patches of depths greater than 7 mm are at most danger of failure due to pressure fluctuations.

(2) Spun cast iron pipes are more vulnerable than pit cast iron pipes. Although spun cast iron pipes are stronger, their thinner pipe walls makes them more susceptible to corrosion and fatigue loading than pit cast iron pipes.

(3) Ageing cast iron pipes with reduced wall thickness and cracks near the critical length could extend their operational life if pressure fluctuations are reduced. Otherwise,

a few ten of thousand of cycles can imminently break a deteriorated pipe. This statement, however, does not consider corrosion fatigue or hydrogen embrittlement which could further reduce the number of required cycles to failure.

(4) Reducing the magnitude of pressure fluctuations is more impactful than reducing mean pressure to decrease the impact of pressure-induced fatigue. A simultaneous reduction of both pressure components should be preferred, as such combined reduction provide a larger synergistic impact.

(5) This analysis motivates that the implementation of pressure reduction schemes, which could reduce both pressure fluctuations and mean pressure, could decrease the occurrence of pipe breaks and extend the life expectancy of ageing cast iron pipes, which can have a major impact on investment decisions.

The conclusions presented in this chapter offer robust theoretical support for the impact of pressure fluctuations on pipe breaks. This, in turn, justifies the use of metrics such as cumulative pressure-induced stress to characterise the effects of pressure transients. The next chapter focuses on the analysis and interpretation of this metric using a machine learning approach. This approach provides an enhanced framework, offering significant improvements over the linear models previously employed in the literature.

Chapter 5

Effect of CPIS on pipe break prediction models

Related publication: Konstantinou, C., Jara-Arriagada, C, & Stoianov, I. (2024) Investigating the Impact of Cumulative Pressure-Induced Stress on Machine Learning Models for Pipe Breaks. *Water Resources Management*. 38. doi:10.1007/s11269-023-03687-7

The Random Forest models used for analysis in this chapter were developed with the assistance of Dr Charalampos Konstantinou.

5.1 Introduction

The previous chapter showed that cyclic loading caused by pressure transients has a clear theoretical impact on the fatigue failure of water distribution pipes. The findings support the application of the cumulative pressure-induced stress (CPIS) concept for the general evaluation of pipes within water distribution networks (WDNs). CPIS was initially explored for its capacity to predict pipe breaks in the work of Rezaei (2017), where a positive correlation between CPIS and pipe breaks was observed. However, it is suggested that the linear analysis used in Rezaei (2017)'s study may have limited

the accuracy of the CPIS model and its interpretation. Additionally, as discussed in the literature review, the current formulation of CPIS has a fundamental issue in how it represents the dynamic pressure (DP) variable, which is not aligned with S-N curve behaviour. Despite this, Rezaei (2017) still identified a relationship between CPIS and pipe breaks. Moreover, the fatigue experienced by water pipes using fracture mechanics progress more rapidly than predicted by S-N curves, suggesting the assumptions of the DP variable reasonably conservative, but requiring also further investigation.

In this chapter, the data used for the CPIS analysis were already preprocessed from the high-speed pressure transducers. This limited the ability to explore new forms of the DP variable in detail. However, further analysis of the original CPIS model was still carried out, using the DP variable as the main variable for explaining pressure fluctuations. A machine learning approach was applied to support this analysis, providing a useful benchmark for any future research that aims to improve the CPIS concept. The key contribution of this chapter is the use of this machine learning framework to systematically study the connection between the original CPIS formulation and pipe breaks.

Acknowledging the role of water pressure in these breaks, it is hypothesised that the relevance of CPIS should increase as pressure-related factors, such as average and fluctuating water pressures play a more significant role in causing pipe breaks, regardless of the specific characteristics of pipes and networks. Under these circumstances, CPIS might effectively replace other pressure indicators. This hypothesis is a central focus of this chapter. Additionally, the chapter utilises two extensive datasets of WDNs in the UK. This method offers improved flexibility in capturing complex, non-linear processes, a step beyond traditional methods for evaluating CPIS. Moreover, random forest modelling allows for the determination of crucial importance metrics, enhancing our ability to understand the relevance and interactions of variables within these models.

The study utilised one dataset (NetA) specifically designed to evaluate the ability of CPIS to capture information from other pressure-related metrics. Subsequently, a second, larger dataset (NetB) was employed to assess the impact of CPIS on model performance.

These analyses provide valuable insights into the role of water pressure fluctuations in pipe breaks and highlight the necessity for developing performance metrics to monitor and evaluate water pressure dynamics. Fundamentally, this chapter provides a machine learning framework to aid in the interpretation and assessment of CPIS and pipe breaks, thereby providing water management professionals with critical information and knowledge to justify continuous monitoring and assessment of water pressure fluctuations.

5.2 Methodology

Machine learning models, specifically random forest models, are applied on two comprehensive datasets to evaluate the power of CPIS to infer pipe breaks. An overview of the proposed methodology is presented in Figure 5.1.

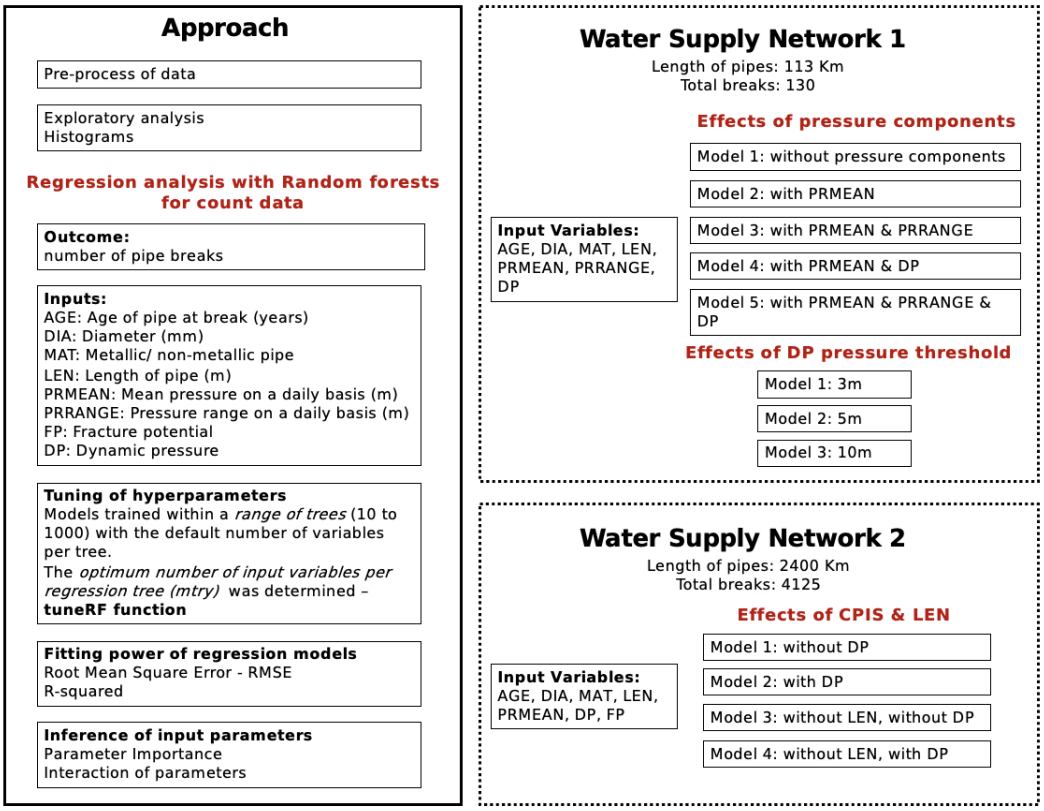


Figure 5.1: The approach followed in the two WDSs. From: Konstantinou et al. (2024).

5.2.1 Data curation and exploratory analysis

Two datasets, NetA and NetB, of historic pipe breaks and associated pipe properties have been provided by two water utilities. The data collected include environmental, operational and pipe intrinsic factors. Details of these factors are shown in Figure 5.1. NetA is characterised by its detailed and good quality data, and the primary focus for its use was to compare the explanatory power of CPIS against other pressure metrics. NetA was also used to assess the impact of the cycle amplitude threshold (minimum cycle amplitude loading) in the explanatory power of the CPIS. The insights gained from NetA were then utilised to collect more targeted information from NetB. Furthermore, NetB is notably larger in scale compared to NetA. This approach allows for a comprehensive analysis that benefits from the strengths of both datasets.

High-resolution pressure was collected from high-frequency pressure monitoring devices over a period of 2-3 weeks for each network. The water pressure variables used in the analysis were obtained from measurements recorded by InflowSense™ pressure monitoring devices. These devices were strategically positioned at chosen locations within the network to ensure adequate spatial coverage. Both the mean pressure and dynamic pressure were preprocessed before they were available for this investigation. The mean pressure data was utilised to calibrate a hydraulic model, enabling the estimation of mean pressure values in all pipes throughout the network. The dynamic pressure metric (DP) was extrapolated throughout the network with an energy dissipation linear model from a past research (Hoskins & Stoianov, 2017; Rezaei, 2017).

5.2.2 Random forest regression and variable performance

In this chapter, the dependant variable provided for analysis was the number of pipe breaks per pipe. For this reason, a random forest algorithm, adjusted for count data (Liaw & Wiener, 2002) was selected instead of the binary logistic regression previously used in Chapter 3. This algorithm selection was also recommended by the work of Konstantinou

& Stoianov (2020). In their study, the dependent variable of the models was the number of breaks per pipe segment. The study suggested that the algorithm offered superior fitting capabilities compared to other machine learning algorithms (Konstantinou & Stoianov, 2020).

In random forest (RF), a forest consists of multiple decision trees and each tree is built with a random subset of covariates (Liu et al., 2010). The conventional regression tree tends to overfit the input data set (Sadler et al., 2018), while RF overcomes this problem via its randomness feature. According to the algorithm developers (Breiman, 2001; Breiman & Cutler, 2004), there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error since it is estimated internally, during the run with the use of the out-of-bag error (OOB). The OOB error is a unique feature of RF that uses the samples that were not used in training a particular tree in the forest to measure the error (roughly one-third of the cases) and has been shown to be impartial in numerous testing.

RF has two main hyper-parameters (see Figure 5.1), the number of trees to grow in the forest (range of trees), and the optimum number of predictor variables randomly sampled as candidates per regression tree ('mtry'). A sensitivity study was conducted to select the appropriate hyper-parameters. The models were first trained within a range of trees (10 to 1000) with the default number of variables per tree (one-third of the variables, or max number of variables). Then, the optimum number of predictor variables per regression tree was evaluated using the 'tuneRF' function in the 'randomForest' package in R (Liaw & Wiener, 2002). The function develops RF models using a range of 'mtry' parameters, one at a time, and determines the appropriate value by minimising the the OBB error within the Random Forest. A 5-fold cross validation was also conducted for hyper-parameter tuning to confirm the values of mtry and ntree, yielding the same results.

The models were fitted using the entire dataset with the objective of conducting a causality analysis on the historical data. The goal was to identify cause-and-effect relationships between the input variables, assess the strength of these relationships, and explore interactions among different variables. The emphasis was on learning from the current dataset

rather than making predictions. The use of OOB, enables the model to be trained, fitted, and validated simultaneously, while mitigating the risk of overfitting. This approach helps ensuring the reliability and generalisation of the model's performance (Hastie et al., 2017).

The count models' performance was assessed using the R-squared and root mean squared error (RMSE) metrics, derived from the comparison of predicted and actual break values. These metrics are commonly employed to evaluate the accuracy of statistical regression models, as shown in Figure 5.1. Following the model development, the contribution of each feature to the model was calculated by determining its contribution in each tree of the model. This calculation involved measuring the total decrease in node impurities resulting from splitting on the variable averaged over all trees (Gini index). The interactions between covariates were also of interest especially for the pressure components and were also evaluated. The interactions between two variables are defined via the prediction function decomposition and are measured by Friedman's H-statistic (Molnar, 2018).

5.3 Case study – Description of the networks

Two datasets (NetA and NetB) were prepared and utilised for the analysis. Both datasets have the total number of breaks experienced by a pipe as the dependent variable. Apart from the water pressure variables, which were derived from high-frequency pressure monitoring, all other variables were obtained from the records maintained by the water utility. The investigations conducted on each dataset are depicted in Figure 5.1. NetA served as a 'training' case study, where a thorough analysis was conducted to offer insights and recommendations on defining the CPIS and its relationship with variables representing pressure components. Different thresholds of event amplitudes for the determination of DP have been also used to assess their effects on the regression fit performance. Following the conclusions drawn from the analysis on NetA, NetB was utilised as a 'validation' case study.

5.3.1 NetA description

The total length of pipes is about 113 Km. A total of 2501 entries was obtained with 100 entries representing pipe segments with breaks (4% of all observations - each record might include more than one break). The information obtained regarding the breaks contains the number of breaks for each segment. The pipe age falls in the range between 0 and 100 years. The pipe segments are divided such that most of the lengths are about 50 m (Figure 5.2 (a)). In this network, most of the material is metallic accounting for approximately 70% of the total pipes and the non-metallic material is mainly plastic. Both the mean pressure and the pressure variations have very similar values in all pipe segments and are not distributed evenly. The mean pressure is around 50 m while the pressure range is around 5 to 10 m. The DP values are concentrated around 500 (Figure 5.2).

5.3.2 NetB description

The second network consists of a larger sample containing approximately 79,750 observations with 977 pipe break records (each record might include more than one break). The failures represent only 1.2% of the total documented observations. The total pipe length for this set is 2400 km. Most of the pipe segments have length of 0-50 m and the average diameter is 80-120 mm (Figure 5.2). Most of the pipe material in this area is asbestos cement, however, the metallic and non-metallic pipes are almost equal. The histograms of the pressure components show that the mean pressure follows again a normal distribution and the DP has significantly larger values compared to NetA.

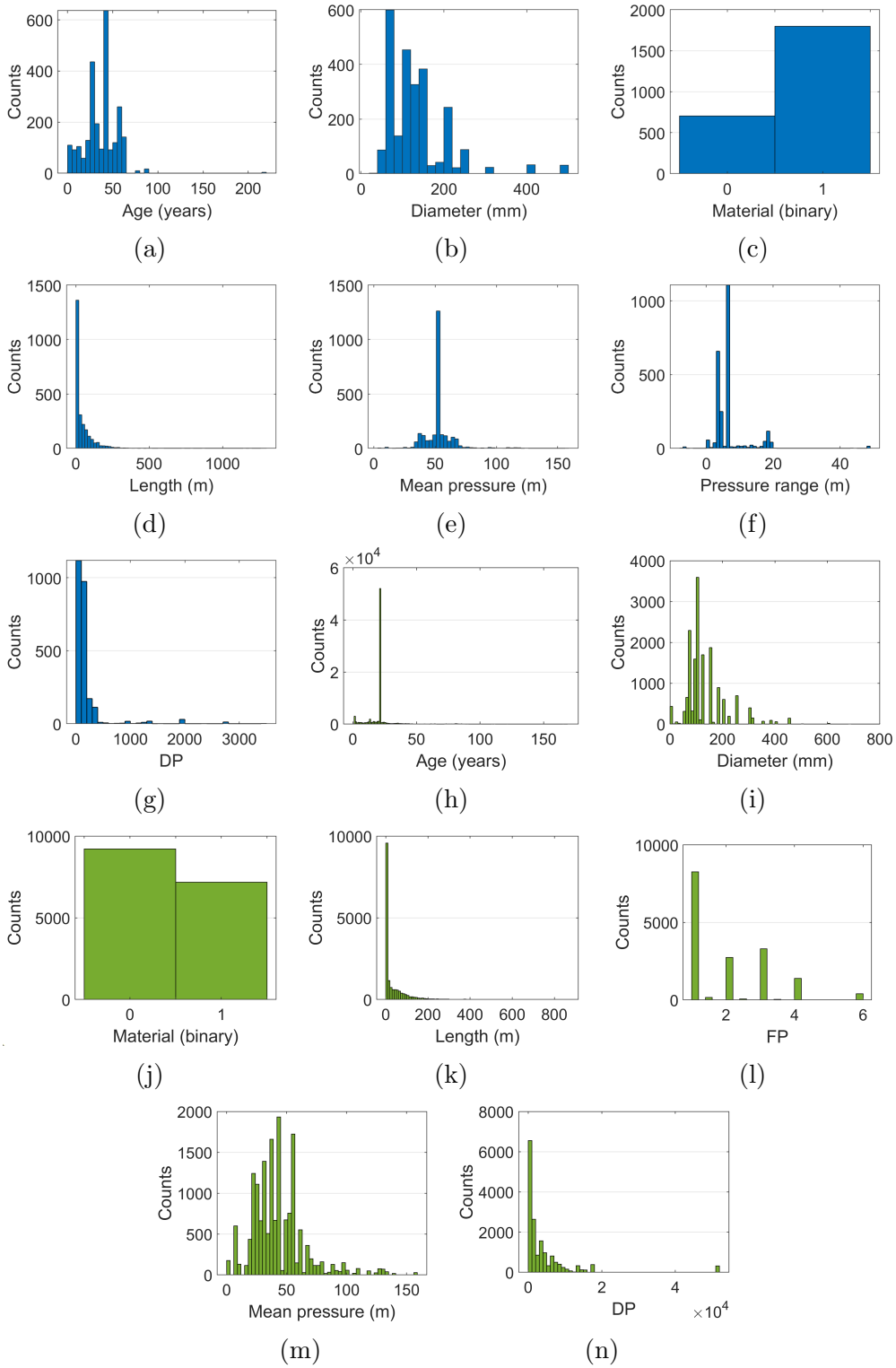


Figure 5.2: Classification of network characteristics for (a-g) NetA and (h-n) NetB. From: Konstantinou et al. (2024).

5.4 Results and discussion

5.4.1 NetA

In NetA, the significance of CPIS is assessed after determining the role of the pressure components. Part of the analysis is the parameters' importance and the interaction terms strength which is calculated with the Friedman's H-statistic (Molnar, 2018) including the pipe breaks based on the developed regression fit.

Investigation of pressure components effects

First, the overall role of the pressure components is assessed by fitting RFs with and without pressure components. Once the level of influence is determined, the PRRANGE and DP effects are compared. These are 'similar' metrics, however, DP carries more information (the event cycles). Since the sample size is smaller in this network compared to NetB, the number of breaks is also smaller. All the pipes experience 0 to 4 breaks with one pipe segment only experiencing 11 pipe breaks. The outlier of 11 breaks was not removed from the sample as the random forest approach appropriately deals with extreme values. Models were developed for the following combinations of pressure components:

- Model 1: without pressure components
- Model 2: with PRMEAN
- Model 3: with PRMEAN and PRRANGE
- Model 4: with PRMEAN and DP (CPIS)
- Model 5: with PRMEAN, PRRANGE, and DP

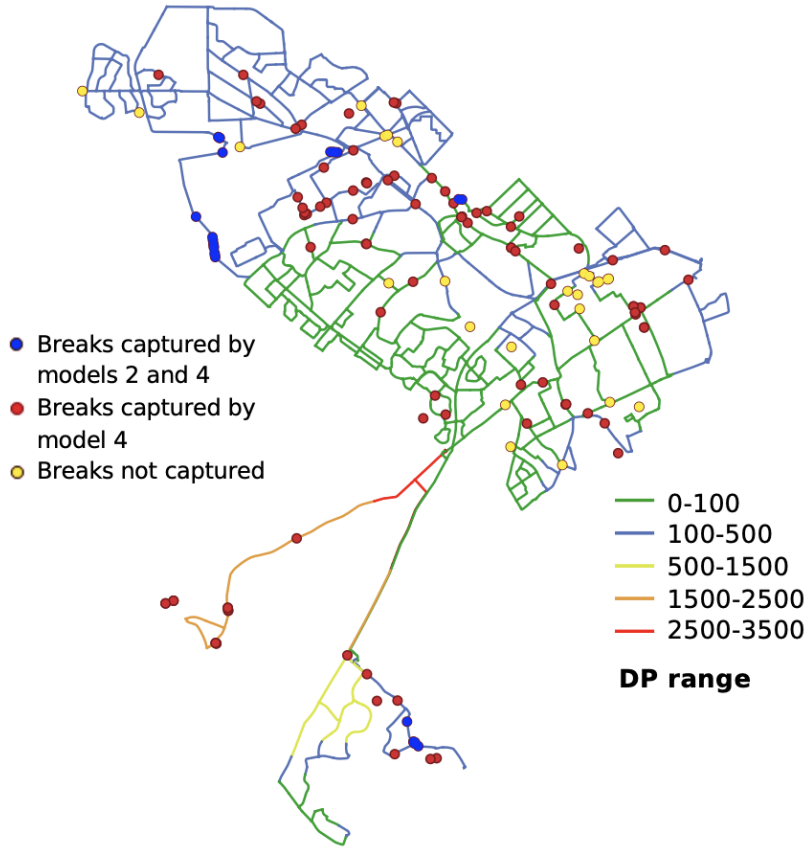
Models 1 and 2 are compared to demonstrate the impact of different pressure metrics in the system. Models 3 and 4 are of particular interest as DP effectively replaces PRRANGE

and includes additional information such as event cycles associated with amplitudes. Model 5, on the other hand, is not a realistic scenario as DP includes the information present in PRRANGE. However, it is included to determine if any information loss occurs when replacing PRRANGE with DP.

The models were assessed based on the R-squared and RMSE values from the plots of predicted breaks versus actual breaks. In the model without pressure components the R-squared value was 0.56 (RMSE of 0.2805) and became 0.62 when PRMEAN was included (RMSE of 0.2708). Therefore, the loading components of this network are important as the model becomes significantly more accurate once the first pressure component of mean pressure is added. This is the first operational factor that is included in the regression analysis beyond the environmental and network characteristics factors. The combination of PRMEAN and PRRANGE increases even more the accuracy of the fitted models to reach an R-squared value of 0.76. Equivalently, the root mean squared error decreases from 0.2708 to 0.2092 once PRRANGE is added in the model as a covariate (22.7% percent decrease in error). However, PRMEAN and DP (CPIS) provide by far the best fits (R-squared of 0.81 and RMSE of 0.2002); 30.5% increase in accuracy compared to Model 2 and 6.6% increase in accuracy compared to Model 3 (comparison in terms of R-squared values). The R-squared and RMSE values of all three components (model 5) are the same as in the former case (0.8113 and 0.2011, respectively) and demonstrates that there is no loss of information when removing PRRANGE. In fact, DP is a stronger pressure component as model 5 performs slightly worse compared to model 4.

Figure 5.3a shows a map illustrating the range of DP values within the network. The breaks have been marked also for different cases. The blue dots represent the breaks captured by model 2 (without the addition of DP), the blue and red dots are the pipe breaks captured by model 4 (with the addition of DP) and the yellow dots are the breaks that model 4 failed to capture. The addition of the DP as a covariate increases dramatically the performance of the model. Figure 5.3b shows the confusion matrix of model 2 in which DP was not used whilst, Figure 5.3c presents these findings for model 4 in which DP was added. The pipes that have experienced one or more breaks have been clustered

together. It is clear from the two matrices that the addition of DP helps in improving the estimations related to the pipes that have experienced a break. The false negatives decrease from 88 to 26 and the true positives increase from 9 to 71 providing a significant improvement in the model performance once DP is added in the model.



(a)

Confusion matrix- Model 2		Actual number of breaks	
		0	≥1
Estimated number of breaks	0	2402	88
	≥1	2	9

(b)

Confusion matrix- Model 4		Actual number of breaks	
		0	≥1
Estimated number of breaks	0	2400	26
	≥1	4	71

(c)

Figure 5.3: NetA results. (a) The range of DP values per pipe in NetA. The confusion matrix for (b) model 2 and (c) model 4. From: Konstantinou et al. (2024).

The parameter significance is also examined for each model (Figure 5.4 (a)). Generally, the length is the most important parameter followed by age and pressure components.

The pipe length is not a factor itself that causes a break, but it is used as a variable in statistical models. As shown in the histogram of the pipe length in Figure 5.2, very few pipe segments are large, and these experience a high break rate. It was proved by Konstantinou & Stoianov (2020) that the variable's importance relative order did not change when adding or removing the length as a variable and the results were consistent between the two cases. In other words, length does not affect the functional relationships and importance of factors affecting pipe breaks.

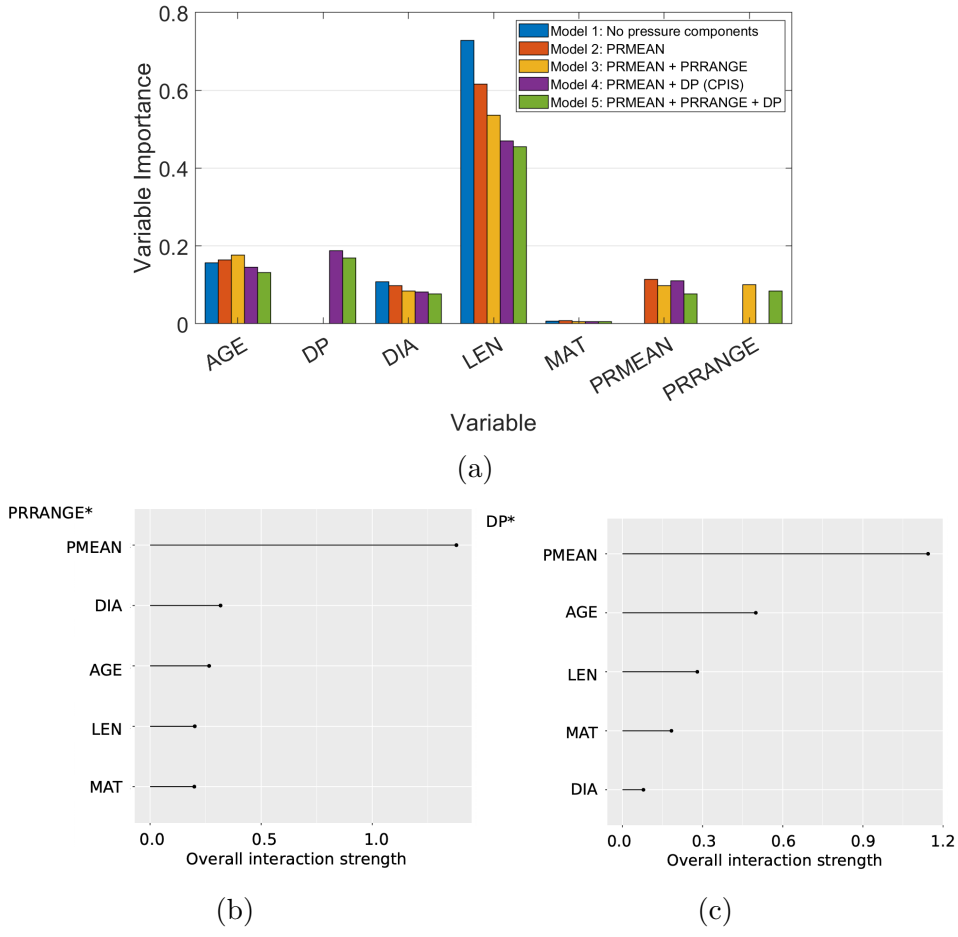


Figure 5.4: Variable importance estimates and interaction strength for NetA. (a) Variable importance for the 5 models for NetA. Overall interaction strength of (b) PRRANGE (model 3) and (c) DP (model 4) with the rest of the variables in the model. From: Konstantinou et al. (2024).

When the PRMEAN variable was added to the model (model 2), it emerged as the third most important covariate, following length and age. Notably, the importance of the length variable decreased significantly with the inclusion of PRMEAN. This suggests

that PRMEAN contributes significantly to the model's predictive power and may have a stronger association with the dependent variable compared to the length variable. The inclusion of the PRMEAN variable likely captured relevant operational aspects that influence the occurrence of pipe breaks. Once the first pressure component was added, which is an operational factor, the significance of length decreases. When PRRANGE was added (model 3), the significance of length decreased even more, and the significance of PRRANGE was similar to PRMEAN. PRMEAN decreased as it contains part of the information of the former variable. When DP was added (model 4), it became the second most significant predictor with length's importance reducing substantially while PRMEAN remained at approximately the same levels. This was expected as amongst the three operating factors (PRMEAN, PRRANGE and DP), the DP contains more detailed information. In model 5, where both DP and PRRANGE were included, the importance of PRMEAN, PRRANGE and DP is affected the most while the rest of the variables exhibited marginal changes. This proves that there is an overlap between the information carried by each of these covariates.

Finally, the strength of the interactions was also assessed for two variables (PRRANGE for model 3 and DP for model 4) with respect to the rest used in the model Figure 5.4 (b-c). The strongest interaction appeared to be between the pressure components in both cases proving that they have multiple and different effects on the models. DP accounts for the number and magnitude of pressure fluctuations experienced by a pipe which produce crack propagation. The mean pressure influences the rate at which crack propagation occurs. Fatigue-related failures are linked to the age, length and material type which explains the rest of the interactions. The rest of the interactions were combinations of pressure components and pipe characteristics. For model 3, the PRRANGE interacts to a lesser extent with DIA, AGE, LEN and MAT. On the other hand, for model 4, DP shows a high interaction with AGE and LEN. This explains why LEN's importance reduces from 0.75 to 0.45 (a 40% reduction) when DP is added to the model as shown in Figure 5.4 (a).

The effects of the DP predictors calculated at different thresholds

The DP metric was calculated in three different ways for NetA. The first calculation was for cycles of amplitude at least 5 mH_2O (DP) which was shown before, the second for cycles of amplitude of at least 3 mH_2O (DP_3), and the third for cycles of amplitude of at least 10 mH_2O (DP_10). The calculated metrics show that as the threshold decreases, more information is retained in the DP metric. More of the short events amplitude events which are more frequent are captured. The threshold of 10 m excludes a vast amount of information.

The parameter importance analysis showed that as the DP becomes more detailed, its significance increases (approximately 16% for DP_10 and 18% for DP_5 and DP_3). However, the accuracy of the models remains approximately at the same levels for 3 mH_2O and 5 mH_2O indicating that lower thresholds below 5mH_2O do not add significant additional information. For 10 mH_2O, the accuracy of the model starts to decrease because it captures only the high amplitude events which are related to a lesser extent to the fatigue-related failures as these are usually high frequency events with low amplitudes. Therefore, it is not suggested to deviate much from the lower range. Between the three values, a threshold value of 5 mH_2O is reasonable to be chosen according to this investigation.

5.4.2 Development of models for NetB

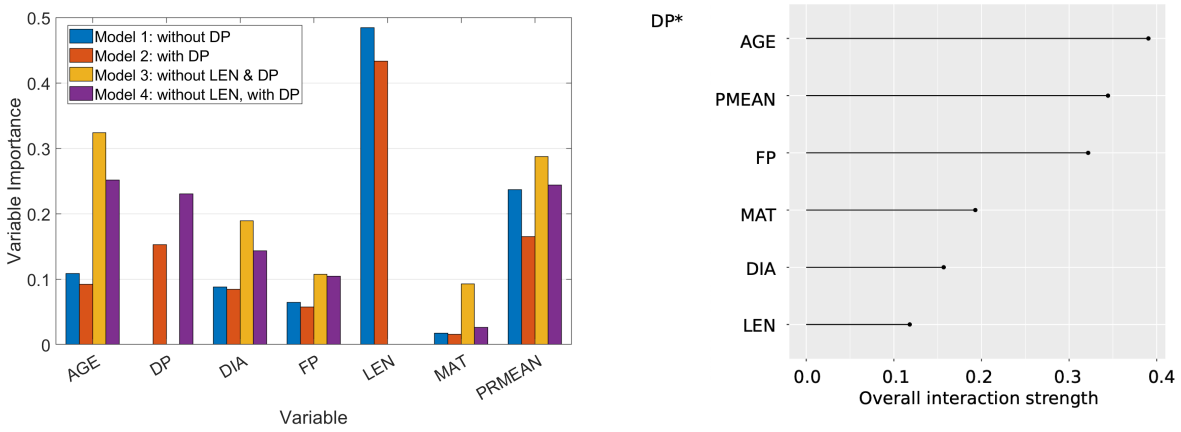
The information obtained from the analysis of NetA is used to validate the hypothesis that CPIS is an important covariate in another WDN with completely different characteristics. The focus in NetB is shifting to identifying the performance of models developed in explaining pipe breaks with and without the CPIS metric (with and without the addition of DP and mean pressure). This is also the ultimate recommendation of this study, the assessment of whether an operator needs to shift their attention to the dynamic components of the network and make plans on how to proceed with regulating the mean pressure and dynamic pressure.

The NetB dataset is much larger in comparison to NetA and is used in order to examine the sample size effects and the role of CPIS in a completely different network with very different environmental and network characteristics. Regression models were fitted with and without DP to assess the role of DP on this network, while the variable LEN has been also examined (models were developed with and without LEN) since it was shown previously to be an important parameter. Models were developed for the following cases:

- Model 1: without DP
- Model 2: with DP
- Model 3: without LEN and DP
- Model 4: without LEN, with DP

The performance of the model increased quite substantially once the DP variable was added as the R-squared value increased from 0.68 to 0.70 (3% increase) and RMSE decrease from 0.2144 to 0.2088 (2.6% decrease in error). The model developed without the addition DP shows that the most important covariate was LEN followed by PRMEAN. The results are almost identical with the case of NetA, which builds confidence about the results. When DP was added it becomes the third most important covariate with an importance percentage close to that of PRMEAN which was the second most important covariate. LEN's and PRMEAN's importance decreased (Figure 5.5a).

Very similar outcomes were derived when developing models 3 and 4 (without adding LEN as a covariate). When LEN was not included as a variable DP, PRMEAN and AGE became the three most important covariates. All input covariates' importance increased proportionally without their relative ranking changing when LEN was removed from the input variables. The interaction strength of DP with respect to the other variables (Figure 5.5b) demonstrates strong interactions between DP and age, DP and PRMEAN, and DP and FP. This comes in contrast to the results of NetA, however, this system is exposed to more aggressive environments with higher fracture potential (FP) when the levels of



(a) (b)

Figure 5.5: Variable importance estimates and interaction strength for NetB.(a) Variable importance for NetB (b) Interaction strength between DP and the rest of the variables. From: Konstantinou et al. (2024).

pipe age and internal pressure components are higher. This finding is in agreement with the fact that DP values in NetB were considerably higher compared to NetA. Figure 5.5b also shows the complexity level of this system which is much higher compared to NetA.

5.4.3 Assessment of the impact of the cumulative pressure-induced stress as a metric in pipe break models

In both networks, there was no single dominant factor that was solely responsible for pipe breaks. However, among the three groups of factors (operational, environmental, and network characteristics) contributing to pipe breaks, the loading components group emerged as the most important. The size of the dataset did not have any impact on the model outcomes. In fact, the importance of the CPIS in both NetA and NetB remained relatively similar despite differences in sample size and network characteristics. This suggests that the predictive power of the model for CPIS is robust and consistent across datasets with varying sizes and network characteristics.

Based on the analysis, when loading factors such as mean pressure, maximum pressure, and pressure variations are found to be influential, the CPIS also becomes important. Correlation tests indicate that DP is not highly correlated with the other variables, sug-

gesting that it captures effects that are not accounted for by any other covariates. As a result, DP effectively replaces all other pressure components except mean pressure. Consequently, the CPIS, defined by PRMEAN and DP, serves as a comprehensive metric for capturing these effects.

The relationship between pipe breaks and the various input variables is complex and highly non-linear. Consequently, the proposed framework, which utilises machine learning techniques to capture these relationships, provides a fast and efficient method to guide industry operators in making informed decisions regarding their pressure management strategies. The framework's output is a recommendation regarding the criticality of PRMEAN and DP variables. These recommendations can help industry operators prioritise and focus their efforts on managing and controlling these variables effectively. By understanding the criticality of these variables, operators can implement tailored pressure management schemes to reduce the likelihood of pipe breaks and optimise the overall performance and resilience of the WDNs.

Random forests are a powerful tool for pattern recognition, though, the algorithm's predictions are limited to the range of the input dataset. This limitation can pose challenges when encountering scenarios where a static pressure component in the future exceeds the historical mean pressure values.

Despite its limitations, this research highlights the importance of pressure management control in WDNs. It represents a crucial step towards conducting field trials or experiments. It is recommended to closely monitor a cluster of pipes within a population that exhibits characteristics representative of the entire network. By continuously and rigorously monitoring pressures at high frequency rates, various features such as real-time fault or anomaly detection can be detected. Early warnings can also be provided during the initial stages of an event. In the wear-out phase of the material, the frequency and amplitude of the events become prominent factors contributing to the degradation process. By understanding and monitoring these factors, proactive maintenance and management strategies can be implemented to mitigate pipe failures and extend the lifespan of the

network.

5.5 Conclusion

This chapter investigated how various hydraulic loading components affect pipe breaks, specifically concentrating on dynamic components as measured by the cumulative pressure-induced stress metric (CPIS). Random forest regressors were used to determine the significance of CPIS as a covariate for modelling pipe breaks in two water distribution networks. The results showed enhanced performance across all models when dynamic pressure (DP) was incorporated, indicated by a reduction in RMSE. DP more efficiently represents the dynamic effects compared to the pressure range variable (PRRANGE), which merely suggests the possible spectrum of pressures. Furthermore, DP not only supplants PRRANGE as a covariate but also contributes extra information to the models. Therefore, the CPIS metric (PRMEAN and DP) was found to be a significant factor responsible for pipe breaks.

While the importance of adaptive pressure management is widely recognised in the water industry, the relationship between input variables and pipe breaks is notably non-linear, adding complexity to any proposed solution. Consequently, the presented framework provides an effective method for aiding industry professionals in making informed decisions about pressure management tactics. The outcomes of the framework include guidance on the relevance of PRMEAN and DP variables. This chapter demonstrates the significance of ongoing monitoring and systematic gathering of high-resolution pressure data as integral elements of a comprehensive asset management strategy for water distribution networks.

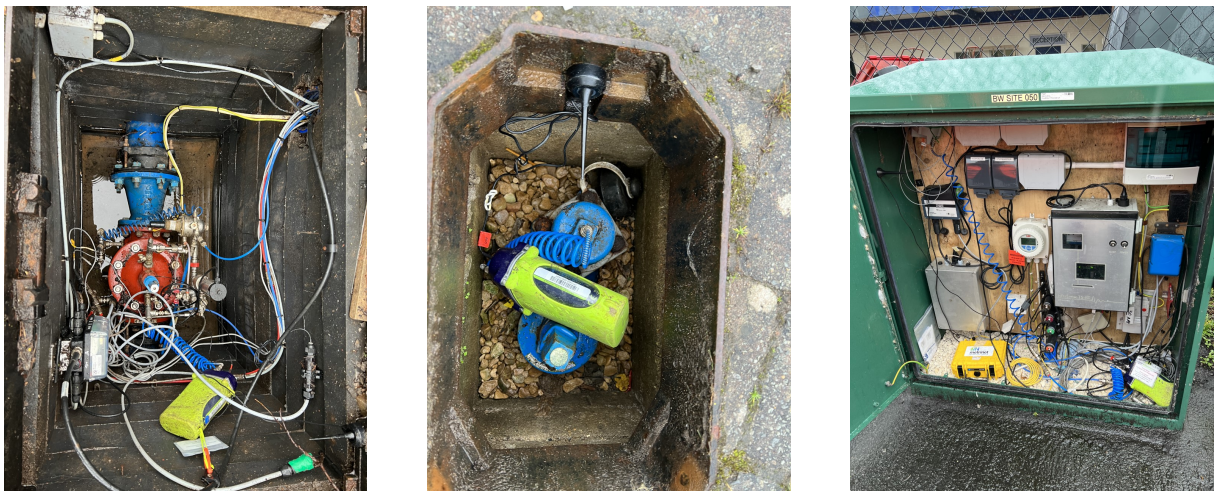
Chapter 6

High resolution pressure monitoring for CPIS estimation

Related publication: Jara-Arriagada, C, & Stoianov, I. (2022) High resolution water pressure monitoring for the assessment of fatigue damage in water distribution pipes. *Proceedings of the 2nd International Joint Conference on Water Distribution System Analysis (WDSA) & Computing and Control in the Water Industry (CCWI)*. 1220-1235. doi:10.4995/WDSA-CCWI2022.2022.14813

6.1 Introduction

The cumulative pressure-induced stress (CPIS) has been investigated in earlier chapters from motivational perspectives, causal mechanisms, and suitability for pipe break prediction models. These previous investigation support further analysis of CPIS through large-scale experiments. The primary goal of such experiments is to give water operators confidence in using CPIS for managing pipe breaks. To conduct these experiments effectively, several technological requirements must be met to ensure that CPIS can be applied in long-term operations, beyond monitoring trials.



(a) (b) (c)

Figure 6.1: Installation of pressure transducers in a) pressure reducing valve (PRV) b) fire hydrant c) telemetry kiosk. From: Jara-Arriagada & Stoianov (2022).

This chapter focuses on determining the optimal sampling resolutions and pressure monitoring settings needed for the automatic calculation and implementation of CPIS in large-scale experiments. Two years of high-frequency pressure monitoring data are analysed for this purpose. Additionally, histograms are used to visually compare dynamic pressure profiles with calm pressure profiles from the available data.

6.2 Methods

6.2.1 Pressure monitoring technologies

The data used in this study was acquired utilising high frequency pressure monitoring devices (Hoskins & Stoianov, 2014) installed in various locations. The sampling rate was 128 S/s (Samples per second), although some of the data was only available for analysis at 1 S/s. This is because the utilised battery operated devices were setup to send 1 S/s sub-sampled data from 128S/s locally stored data. The locations for the installation of the pressure monitoring devices were pressure reducing valves (PRVs), fire hydrants, and telemetry bollards/kiosks. Example illustrations of the installation locations and the pressure monitored devices are presented in Figure 6.1.

6.2.2 Cycle counting

Measuring water pressure at high sampling rates unveil loading conditions in pipes otherwise undetectable at lower frequencies. Figure 6.2 contrasts the difference between water pressure measured at 1 sample every 15 min for common SCADA systems versus a high speed logger at a rate of 1 S/s. The water pressure profile presented from the high speed logger shows clear pressure variations at different magnitudes. From a mechanistic point of view, a single load of such magnitude does not pose any threat to a pipe. However, as presented in Chapter 4, several repetitions of cyclic loadings can propagate cracks up to a point at which the pipe can become vulnerable to normal operating loadings.

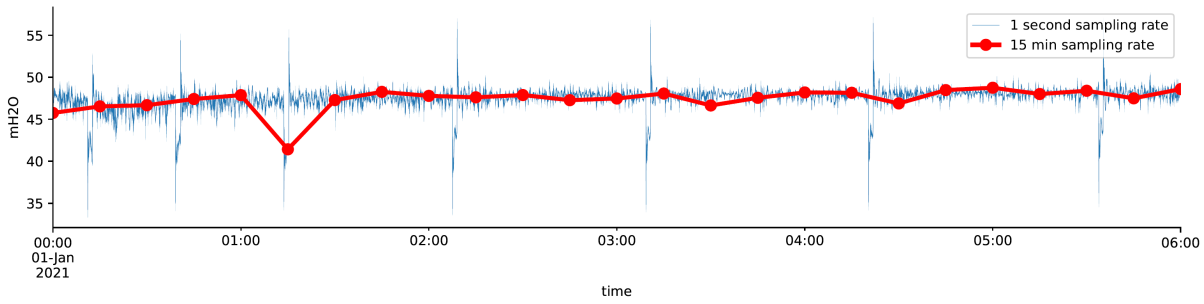


Figure 6.2: Comparison of pressure profiles obtained by a low and high frequency sampling rates. From: Jara-Arriagada & Stoianov (2022).

To assess the propagation of cracks using fracture mechanics (Chapter 4), the information of the magnitude and number of cyclic loadings presented in the load history must be extracted. Similarly, to estimate the DP variable for CPIS, the same information must be extracted. This task is not trivial. In contrast to constant amplitude loadings, where the amplitudes of each cycle are well defined, there is no precise definition of what constitute a cycle under random load histories (Cheng & Broz, 1986). In such cases, cycle counting algorithms that perform heuristics to count cycles must be utilised (ASTM, 2017).

Several techniques for such cycle counting exist, and the results can vary between one or other technique (ASTM, 2017). In the field of fatigue analysis there is a preference for the rainflow cycle counting technique (Rychlik, 1993). The preference of this algorithm relies on its capability to count cycles within cycles (Figure 6.3), and the possibility to

relate its results with stress and strain hysteresis curves (Anzai & Endo, 1979). It has been also empirically demonstrated that the use of this algorithm produces satisfactory loadings for prediction of actual fatigue life (Dowling, 1972). Given its preference and general performance acceptability, this chapter implements this algorithm in the version of the ASTM E1049-85 (ASTM, 2017) for cycle counting of the time series pressure data.

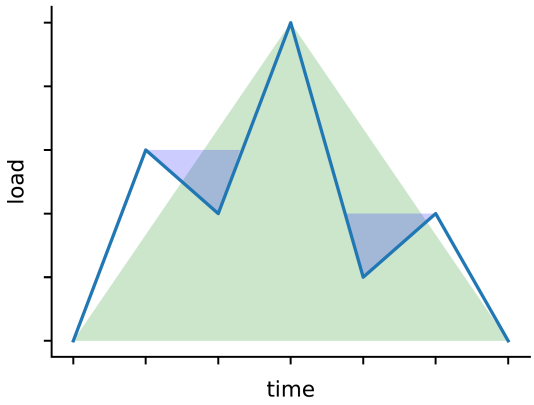


Figure 6.3: Visualisation of cycle detection by the rainflow cycle counting technique. From: Jara-Arriagada & Stoianov (2022).

The outcome of the rainflow cycle counting technique is a distribution or histogram that provides the amplitude of a cyclic load and the number of repetitions for each cycle. This information is, however, not provided in sequence, which means that information of history effects are missing, i.e., the occurrence of high load before a small one. Nevertheless, this information is usually required in detailed fatigue analysis performed "cycle by cycle" where crack closure effects can be considerable at low R ratios (Anderson, 2017) (The R ratio is the ratio between minimum and maximum pressure). For the majority of cases where R ratios are high, a simple average of the cycle counts for a specific load history can provide good enough estimates for fatigue analysis (Anderson, 2017).

6.2.3 Assessment of the required sampling rate

Some pressure monitoring devices were set to communicate their raw data of 128 S/s, while others only communicated a sub-set of 1 S/s derived from the continuously acquired 128 S/s. Pressure data acquired at 128 S/s have been utilised to localise sources

of transients. Pressure transients in WDNs are the consequence of travelling waves (Misiunas, Vítkovský, Olsson, Simpson, Asce & Lambert, 2005). These waves are generated due to sudden changes in flow conditions, and can reach velocities up to 1200 m/s in rigid pipes (Rathnayaka et al., 2015). The optimal sampling rate for measuring such fast propagating waves depends on the purpose of the analysis. If the intention is to fully characterise the shape of a wave and their exact arrival time in space, then a sampling rate of 128 S/s or more is required to achieve errors of less than 10 m (Hoskins, 2015). Nevertheless, if the sole purpose of the analysis is to capture the maximum rising edge of wave in order to count cyclic loadings, then the sampling rate can be reduced without substantial loss of information.

Optimising the sampling rate according to the required use is of critical importance because of the large computational resources needed to process and store billions of data points obtained at high resolution. This chapter assesses the loss of information between counting cyclic loading at data sampled at 128 S/S, 10 S/s and 1 S/s. These three levels separated by one order of magnitude were selected on the basis of practical energy consumption and storage capacity savings. In addition, cycle counting is also evaluated after the application of a noise removal algorithm on the raw data.

6.2.4 Dynamic pressure profile assessment

Given the data available, it was deemed useful for future comparisons to develop a categorisation of the pressure profiles obtained form high frequency pressure monitoring. The rainflow cycle counting algorithm was applied on several time series of water pressure measured over periods of one year. Then, water pressure profiles were categorised within 3 categories of calm, regular and dynamic water pressure according to the magnitude and number of obtained cycle counts. This classification can serve in order to reach targets of pressure reductions according to the implications of the water pressure on specific materials.

6.3 Results

This section presents key results based on the determination of cyclic loadings using the rainflow cycle counting algorithm on time series of high frequency water pressure data.

6.3.1 Sampling rate

The calculation of CPIS requires the utilisation of several months of water pressure monitoring in order to provide sensible results on the fatigue effect of cyclic loadings. Therefore, the trade-off of utilising a sampling rate detailed enough to accurately capture cycles generated by rapidly travelling waves, and a sampling rate that is low enough to allow for savings in computational resources and storage capacity was of critical importance. In order to assess an optimal sampling rate, the difference in cycle counting between sampling data a 128 S/S, 10 S/s and 1 S/s was compared. It was also incorporated the result of counting cycles from the raw data after application of a noise removal algorithm. This analysis was conducted on data sampled in a single location near a valve generating transients during 2018 and 2021. The results of this analysis are presented in Table 6.1 and Table 6.2.

Table 6.1: Cycle counting comparison at different time resolutions, 1 year of data 2018. From: Jara-Arriagada & Stoianov (2022).

Stoke Lane	01/01/2018 count	19/12/2018					
range	128 S/s	1 m noise removal	relative error	10 S/s	relative error	1 S/s	relative error
(0.999, 2.0]	584769	440229	25%	485298	17%	226982	61%
(2.0, 3.0]	40527	31988	21%	32162	21%	19484	52%
(3.0, 4.0]	5179	4563	12%	4493	13%	3468	33%
(4.0, 5.0]	1748	1790	2%	1666	5%	1429	18%
(5.0, 10.0]	1537	1241	19%	1456	5%	1292	16%
(10.0, 20.0]	430	421	2%	413	4%	399	7%
(20.0, 40.0]	179	178	1%	150	16%	146	18%
(40.0, 60.0]	5	5	0%	2	60%	1	80%
cycles over 5 m	2151	1845	14%	2021	6%	1838	15%
cycles over 10 m	614	604	2%	565	8%	546	11%

The pressure at the measured location had moderate fluctuations. As Table 6.1 shows, most cycles are in a range below 2 m, with more than 86% of the presented data. This

Table 6.2: Cycle counting comparison at different time resolutions, 1 year of data 2021. From: Jara-Arriagada & Stoianov (2022).

Stoke Lane	01/01/2021 count	31/12/2021					
Range	128 S/s	1 m noise removal	relative error	10 S/s	relative error	1 S/s	relative error
(0.999, 2.0]	604109	324631	46%	473317	22%	281498	53%
(2.0, 3.0]	84528	48762	42%	65497	23%	38827	54%
(3.0, 4.0]	11406	7139	37%	8844	22%	5368	53%
(4.0, 5.0]	2270	1726	24%	1955	14%	1450	36%
(5.0, 10.0]	1580	1267	20%	1383	12%	1122	29%
(10.0, 20.0]	417	407	2%	392	6%	371	11%
(20.0, 40.0]	45	39	13%	37	18%	33	27%
(40.0, 60.0]	3	3	0%	3	0%	3	0%
Cycles >5 m	2045	1716	16%	1815	11%	1529	25%
Cycles >10 m	465	449	3%	432	7%	407	12%

number can reach almost 99% in case of also counting cycles below 1 m. However, these cycles were not included as they were considered noise. It is observed that the application of a lower sampling rate resolution of 10 S/s can achieve better results closer to the counting of raw data. When applying the lowest sampling rate of 1 S/s, the errors increase considerably for lower cycle ranges. Nevertheless, the errors become acceptable when dealing with cycles over 5 m. Depending on the material, this issue could be a limitation of the 1 S/s sampling rate. Nevertheless for materials that possess higher stress intensity threshold, a sampling rate of 1 S/s seems to be adequate. The results over the two analysed years are consistent.

A point of consideration is that the cycle count presented in the table only shows the count of full cycles. However, there are also residual half cycles. Especially for very large, isolated cycles, it is possible that at lower resolutions, only half of a large cycle might be counted. Nevertheless, these isolated events are unlikely to generate large changes in the overall estimations for fatigue in a fracture mechanics analysis. Unless these events are very large in order to reach the fracture toughness of the material, or if these events are capable of proving retardation or acceleration events, they should not pose a source of large errors.

On the basis of the presented calculations, it is clear that a minimum sampling rate of 10 S/s is closest to the original sampling rate. An important consideration is that in

the analysed profiles, a sampling at a rate of 1 S/s can be accurately enough for cycles above 5 m. For lower cycle magnitude, it seems that there could be a proportional loss of information that could be accounted when estimating cyclic loadings.

For fracture mechanics purposes, on particular tough materials, it might be required only to assess cycles above 5 m, for which a sampling rate of 1 S/s is sufficient in operational WDNs. Furthermore, as presented in the previous chapter, for CPIS calculation, a minimum of 5 m seems to be the limit for appropriate explanatory power for general pipe break analysis. Furthermore, materials with very low stress intensity threshold such as asbestos cements might require a higher precision in counting cycles.

6.3.2 Dynamic vs calm profiles

Several time series of one year of water pressure monitoring were investigated. Given the size of the accumulated data over a period of 1 year, and computational processing time, the pressure profiles were analysed using a sampling rate of 1 S/s. From the analysed profiles, three categories based on the cycle counting results were generated. These categories are illustrated with the example pressure profiles of Figure 6.4.

Based on the application of the cycle counting algorithm on several datasets, the observed pressure profiles were categorised as shown in Figure 6.5. In general, this network does not have a high pressure transient activity as compared to networks with large presence of industrial users and pumps (Chapter 7). Nevertheless, it was still possible to provide three categories with large differences between different areas of the network.

The categorisation is based on the shape of the distribution of cycle count for the measured water pressure. This categorisation updates the one provided by Rezaei (2017), which was based on the shape of the water pressure profiles and the visible apparent number of pressure fluctuations and magnitude. The utilisation of the shape of the cycle counting histogram provides better insights into the possible damage that a cyclic loading can generate, and it is easy to visually evaluate the number of cycles that occur over one-

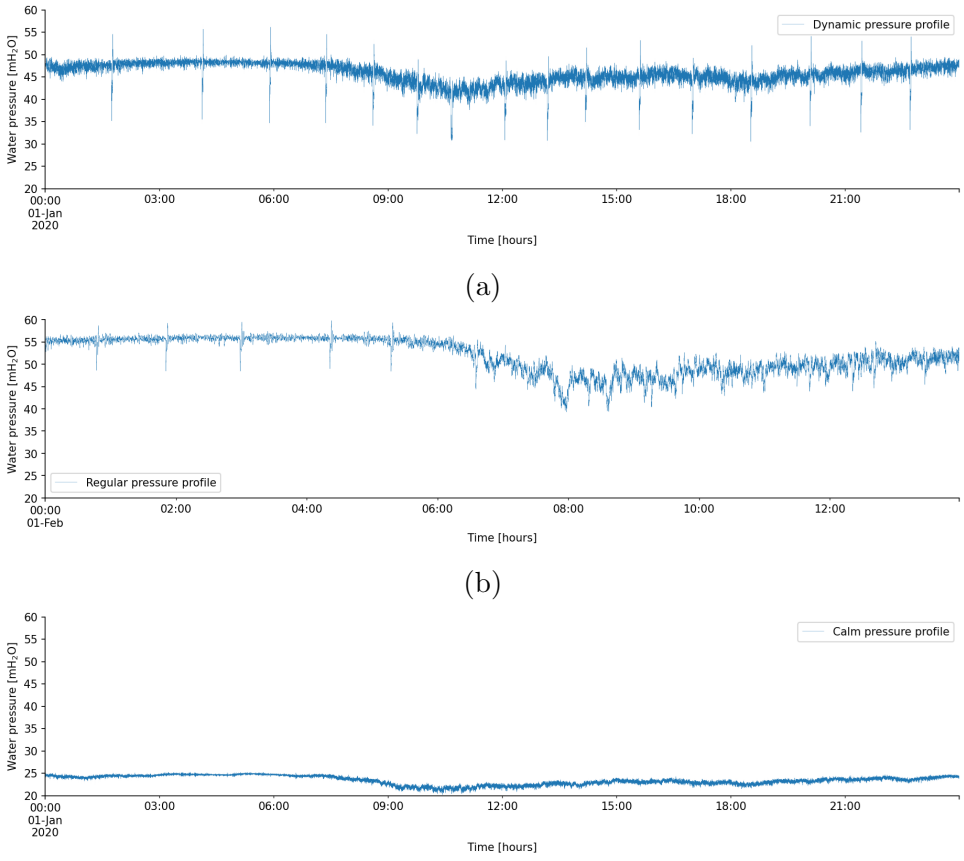


Figure 6.4: Measured water pressure profiles in operational water distribution networks, example of a regular day, a) Dynamic, b) regular, c) calm. From: Jara-Arriagada & Stoianov (2022).

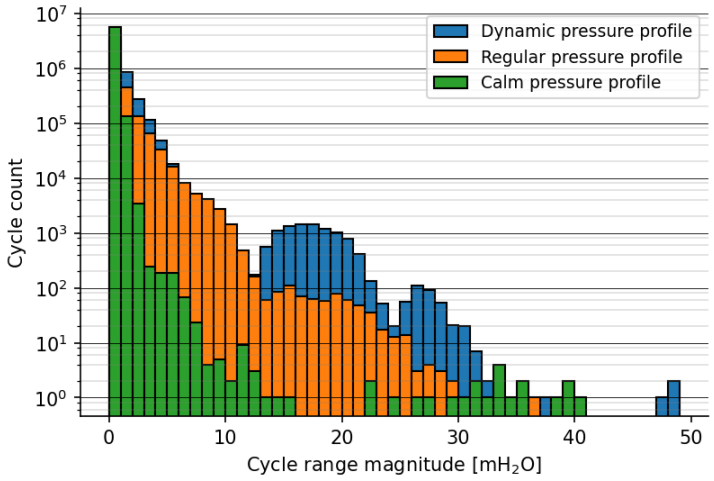


Figure 6.5: Cycle count histogram comparison of dynamic vs calm profiles over one year of pressure monitoring. From: Jara-Arriagada & Stoianov (2022).

year-period.

The presented histogram is in a log scale, which shows an indication of the orders of magnitude of difference between the pressure profiles. These profiles exemplify the results of several locations of the investigated WDN, nevertheless it is also possible to encounter larger profiles at transmission mains and near pressure control devices. For the largest measured pressure profile, more than 1700 cycles with a minimum magnitude of 20 m were counted. This number increase to more than 8000 cycles with a minimum magnitude of 10 m. Over a period of 50 years this can reach up to 400,000 cycles over 10 m. As presented in Chapter 4, such pressure cycles would only be relevant for pipes that are on the brink of collapse, and very poor conditions. However, as indicated, such pipes are not a minority in water networks, and the potential to further reduce pressure activity seems to be of critical importance.

6.4 Conclusion

This chapter has provided an analysis of time series of water pressure obtained from high speed pressure transducers over periods of one year at different locations in a water distribution network. The performed analyses are required in order to optimise the long-term acquisition of high frequency pressure data needed to understand the potential effects of water pressure fluctuations on fatigue failure on pipes through the cumulative pressure-induced stress.

The analysis shows that a sampling rate of 10 S/s is an optimal rate for the sole purpose of counting cycles of water pressure. Nevertheless, 1 S/s can be enough for applications in which cycles over 5 m are of interest. Furthermore, different categories of water pressure profiles were presented on the basis of a proposed categorisation of cycle count histograms generated with water pressure profiles over one year of monitoring.

Although for this particular case study the pressure profiles present a moderate presence of dynamic conditions, the categorisation shows that it is possible to implement reduction of

pressure fluctuations. Such reductions would ensure the operation of quasi-calm networks, which would reduce the impact of fatigue for pipes on the brink of collapse. The findings of this analysis will assist the design of experiments for the assessment of CPIS in large-scale experimental networks bringing confidence in the sampling rates needed for cycle counting. The next chapter will explore another technological challenge for the evaluation of CPIS related to its automated interpolation in networks.

Chapter 7

Interpolation of CPIS across water networks

Submitted manuscript: Jara-Arriagada, C, & Stoianov, I. (2025) Interpolating transient pressure metrics using sensor data for water distribution networks asset management. Under Review. *Water Resources Research*. ISSN: 0043-1397

7.1 Introduction

The utilisation of the cumulative pressure-induced stress (CPIS) concept requires the estimation of pressure cycles for the dynamic pressure (DP) metric for all pipes in a network. This is a complex task given the challenges of modelling pressure transients in water networks. The literature review (Chapter 2) showed that an initial attempt to solve this problem was presented in (Rezaei, 2017), where the DP metric was interpolated using a regression approach. Such approach overlooked a variety of variables involved in the propagation of pressure transients, lacked detailed validation and was used for the analysis of a densely monitored network. However, for a conventional use of the CPIS concept, networks might not be as densely monitored, requiring the investigation of new methods for interpolation of CPIS.

This chapter explores an novel hybrid approach for the interpolation of CPIS that separates the DP metric into components caused by large and repetitive pressure cycles, and small pressure cycles. The hypothesis of this approach is that larger and repetitive cycles, if present, can be identified and modelled within a network, while smaller cycles, which may result from multiple sources, could exhibit spatial patterns over time and therefore be interpolated using spatial interpolation methods commonly employed for geospatial data. This approach is tested using data from two pressure monitoring campaigns conducted in an operational DMA in the UK. These campaigns captured both significant large repetitive transient events, and smaller, unidentified pressure transient readings.

7.2 Methodology

This section describes the proposed methodology for estimating the dynamic pressure (DP) variable throughout a water distribution network. The approach evaluates the determination of DP at unmonitored locations by partitioning the contributions of large, repetitive cycles and smaller, more varied cycles in the recorded data, and then interpolating these components separately. Unmonitored locations correspond to nodes in the network without high resolution transient pressure records. Interpolation methods are routinely employed in geospatial analyses to estimate values that exhibit spatial correlations. Here, it is proposed that the magnitude of dynamic pressure may also display spatially coherent patterns. Furthermore, it is hypothesised that the contribution of large, repeatable pressure cycles to DP is more accurately captured by scaling the output of a pressure transient model.

Under conditions when large pressure cycles occur, transient source localisation techniques can be employed to identify their sources. Once these sources are determined, the propagation and reflection of large pressure cycles can be more accurately simulated through short-duration transient analyses. The results of these simulations are then utilised to establish a threshold value (Section 7.2.3), which in turn facilitates the computation

of two distinct DP metrics at measured locations. The first metric, DP_{high}, is derived exclusively from cycles surpassing the threshold, while the second metric, DP_{low}, encompasses those falling below it. Both metrics are subsequently interpolated, with DP_{high} relying on transient simulation outputs as a reference and also employing spatial interpolation methods for comparison, and DP_{low} being interpolated solely through spatial methods. After estimating the DP metrics for each node in the network, these metrics (DP_{high} and DP_{low} separately or in combination) can be associated to each pipe in the network, for example, by calculating the mean of the DP values of the two nodes connected by a pipe. In this way, DP can be used in pipe break prediction and causality models.

The methodology and underlying assumptions are evaluated using two monitoring datasets obtained from an operational water distribution network in the UK. These datasets, collected under varying conditions and at different sensor locations, provide a basis for validating the proposed approach in a real-world setting.

7.2.1 The DP variable

As presented in the literature review, the mean pressure and the dynamic pressure (DP) metric together constitute the cumulative pressure-induced stress (CPIS), as patented by Hoskins & Stoianov (2017) and further detailed by Rezaei (2017). CPIS was developed to quantitatively represent the influence of pressure transients on pipe breaks, thereby enhancing the predictive capabilities of pipe break models.

The CPIS is a function of the mean pressure (P_{mean}) and the dynamic pressure (DP) as follows:

$$CPIS = f(P_{mean}, DP) \tag{7.1}$$

,where the DP term characterises the contribution of transient pressure fluctuations. It is

computed by summing the products of discrete pressure fluctuation magnitudes (σ_i) and their occurrence counts (n_i) across all observed transient events (M):

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

The mean network pressure is typically estimated using an hydraulic model through extended-period simulations. In contrast, quantifying the DP component of the cumulative pressure-induced stress (CPIS) is more complex, as it necessitates extracting cycle counts and amplitudes from high-temporal-resolution pressure data.

This chapter aims to quantify the DP component of CPIS throughout a water distribution network by employing a practical, reproducible approach that requires minimal computational effort and is specifically tailored to the needs of network operators. As previously discussed, pressure variations captured in high-speed pressure data may be attributed to a limited number of distinct large transient events, as well as numerous smaller transients originating from multiple, less prominent sources (e.g., small commercial users). It is therefore hypothesised that the most accurate estimation of the DP variable is achieved by separately analysing large and small transient events.

7.2.2 Temporal interval for DP calculation

DP is derived by counting cycles of varying ranges in water pressure time series. The calculation interval depends on the intended application. For instance, DP may be calculated daily to continuously monitor pressure fluctuations and facilitate prompt pressure management interventions. Daily DP values are then aggregated into weekly metrics to characterise the network's transient behaviour for pipe break analysis. In this study, DP values are aggregated on a weekly basis to support such analyses; however, the same methodology can be readily applied to daily calculations.

7.2.3 Threshold for DP separation

This chapter proposes introducing a threshold to partition the dynamic pressure (DP) into two components: one composed of large pressure cycles exceeding this threshold and one of smaller cycles below the threshold. The portion consisting of larger cycles (DP_{high}) is to be interpolated using transient model outputs as a reference, and its estimation compared against various interpolation methods. Meanwhile, the portion composed of smaller cycles (DP_{low}) is to be estimated at unmonitored locations solely through spatial interpolation techniques. A graphical representation of this approach is shown in Figure 7.1.

First, if large and repetitive pressure cycles are present, their identification and source localisation is done using transient source localisation techniques, such as the method outlined in Chapter 8. Once the source is identified, a transient model is calibrated to replicate the pressure transient profile and its reflections from a single event. The calibrated model is then employed to simulate the propagation and reflections of the transient wave across the network. For each node in the network, DP values are computed based on the model's outputs, providing a reference for how DP magnitude varies across the network due to the identified large transient source.

These DP values from the transient model are subsequently used to establish a threshold for distinguishing large pressure cycles from smaller ones during DP interpolation. The threshold is determined iteratively, ranging from zero to a value that satisfies two conditions. The first condition requires that DP_{high}, derived from cycles exceeding the threshold at measured nodes, exhibits a strong correlation (close to 1) with DP_{model} values calculated from the transient model at the same nodes. Here, a correlation is considered significant if its p-value is below 0.01, based on a two-sided t-distribution. The second condition established that the absolute percentage change in the correlation coefficient's variation between consecutive threshold iterations is less than 1%. This iterative approach ensures a statistically robust threshold that is close to the optimal value given the data.

Once the threshold is identified, DP_{high} can be interpolated using the transient model

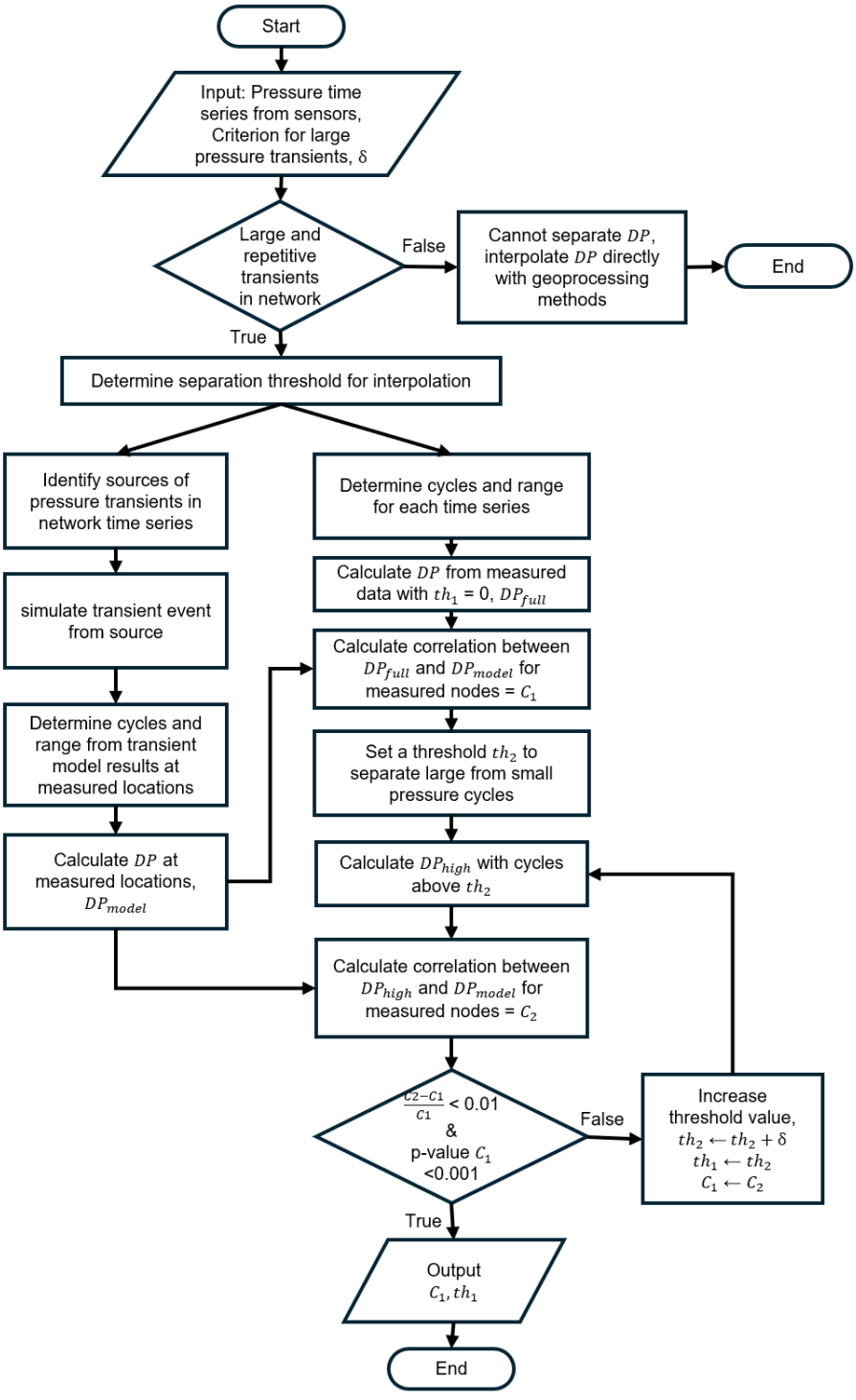


Figure 7.1: Flowchart for threshold determination.

results as a reference and compared against spatial interpolation techniques. The cycles in the measured data with ranges below the threshold are then used to calculate DP_{low}, which is estimated exclusively through spatial interpolation methods. This methodology ensures accurate modelling of the contribution of large cycles to DP while allowing smaller cycles to be approximated through spatial interpolation. Detailed descriptions of the interpolation methods for DP_{high} and DP_{low} are provided in the following sections.

7.2.4 DP above threshold interpolation, DP_{high}

If the threshold determination method is valid, then DP values derived from the transient model (DP_{model}) should exhibit a strong linear relationship with those obtained using DP_{high}. Consequently, the regression line utilised to determine the correlation coefficient between DP_{model} and DP_{high} is expected to capture this linear dependence. By applying this regression line, it becomes possible to estimate DP_{high} at nodes where no direct measurement is available, using the transient model outputs. Furthermore, the performance of various spatial interpolation techniques is compared against this approach, thereby assessing the degree of error introduced by employing interpolation methods for approximating DP.

7.2.5 DP below threshold interpolation, DP_{low}

The DP below the threshold, DP_{low} is calculated using the pressure cycles below the established threshold and above a minimum magnitude. In practice, this minimum magnitude represents the smallest cycle magnitude considered capable of contributing to fatigue-induced failures or other transient phenomena of interest within the system. In this chapter, a minimum pressure magnitude of 0.5 m was selected to explore the feasibility of applying interpolation methods across various types of pressure cycles. Incorporation of this 0.5 m threshold significantly increases the computed DP_{low}, as numerous cycles within the 0.5–2 m range were captured by the pressure monitoring devices. The resulting

DP_{low} at each node is attributed to minor reflections generated by larger transient events, as well as transient occurrences arising from routine consumer usage in residential and small commercial settings, such as restaurants and shops.

7.2.6 Spatial interpolation methods

Various methods for spatial interpolation are available (Li & Heap, 2014), and most rely on the fundamental assumption that values at geographically proximate locations are more similar than those separated by greater distances. Although originally developed for regular grids, such as for precipitation (Lloyd, 2005), wind speed (Ozelkan et al., 2016), and particulate matter concentration datasets (Choi & Chong, 2022), these techniques can be adapted to graph-structured data by using shortest-path distances between nodes. Two widely referenced methods are Inverse Distance Weighting (IDW) (Shepard, 1968) and Kriging (Chilès & Desassis, 2018). IDW leverages an inverse-distance relationship to assign weights, while Kriging incorporates additional spatial information such as spatial correlation and redundancy. However, despite Kriging's more comprehensive utilisation of data characteristics, the existing literature suggests that its performance gain over IDW may be limited when only a few data points are available (Robinson & Metternicht, 2003).

Kriging and its variants rely on characterising the spatial correlation structure between data pairs to estimate values at unobserved locations. This spatial correlation information is typically derived from a variogram, a function that relates half the squared differences in measured values between all possible data point pairs to their corresponding separation distances (see Gringarten & Deutsch (2001) for further details). The shape of the variogram reveals whether spatial correlation is present in the dataset, with smaller squared differences observed between nearby points and larger squared differences typically occurring as the distance between points increases. A well-defined variogram typically displays an increasing trend with distance, indicating the presence of spatial correlation.

For the study in this chapter, variograms were generated to explore the interpolation of DP_{high} and DP_{low} for both datasets (Figure 7.8 in Section 7.4.2). Although DP_{high}

exhibited a slight spatial correlation trend, it was weaker than anticipated for a variable influenced by a single transient-generating process. Meanwhile, the DP_{low} variograms did not show a clear increasing trend, indicating the absence of robust spatial correlation. This lack of correlation may arise from multiple transient sources, each contributing smaller events. Under these circumstances, Kriging is unlikely to yield reliable results. Furthermore, the limited number of available data points may have impacted the robustness of the variogram analysis and the subsequent fitting of spherical or Gaussian models for DP_{high}. Given these observations, it was opted to abandon Kriging-based methods in favour of more direct interpolation techniques that do not rely on spatial correlation. As an alternative reference method alongside IDW, this chapter also employs an interpolation approach based on the Graph Laplacian, a well-documented methodology for estimating missing node data within graph-structured domains.

7.2.7 Inverse distance weighting interpolation

Inverse Distance Weighting (IDW) (Shepard, 1968) is a commonly used method for spatial interpolation of geographical data (Lu & Wong, 2008). The method is based on the principle that the attribute values of two points are inversely related to the distance or connectivity measure between them. When applied to graph data, this distance measure can be defined as the shortest path between nodes (Shiode & Shiode, 2011). To capture non-linear attenuation of attribute values with increasing distance, a power-law decay factor is commonly introduced.

The general formulation of IDW is as follows: The interpolated value z_P at a point P in a two-dimensional Cartesian space, given n known points $D = D_i \in \mathbb{R}^2$ in the neighbourhood of P, each with a feature value $z_i \in \mathbb{R}$, is defined by:

$$z_P = \frac{\sum_{i=1}^{n} z_i w_i(P)}{\sum_{i=1}^{n} w_i(P)} \tag{7.3}$$

where $w_i(P)$ is a weight that depends on the distance $d(P, D_i)$ between P and each known

point D_i, and on an exponent β:

$$w_i(P) = \frac{1}{d(P, D_i)^\beta} \tag{7.4}$$

Numerous variations of this method have been proposed in the literature and integrated into various software packages. In this chapter, the standard IDW formulation was employed, as it exhibited favourable initial performance. An exponent $\beta = 2$ was used, which yielded the best interpolation results (in terms of mean average error) in both datasets. This exponent value is also frequently employed and recommended in studies involving the interpolation of spatial features with limited data (Lu & Wong, 2008; Barudžija et al., 2024).

7.2.8 Graph Laplacian-based interpolation

Interpolation methods based on the graph Laplacian (Garcia Cardona et al., 2022) are another commonly applied method for interpolating georeferenced data, with formulations specifically designed for graph structures (Shi et al., 2017; Narang et al., 2013). In principle, these methods rely on minimising differences in attribute values between connected nodes, thereby producing a smooth transition between observed and interpolated nodes.

The general formulation of the Graph Laplacian-based interpolation approach is as follows: Consider a graph $G = (V, E)$ with vertex set V and edge set E. The Laplacian matrix L is defined as $L = D - A$, where A is the adjacency matrix and D is the diagonal degree matrix. One key property of the Laplacian matrix is the graph Laplacian quadratic form. Let x represent a vector of values associated with every node in the graph. The expression $x^T Lx$ can be expanded as $\sum_{e_{ij} \in E} w_{ij}(x_i - x_j)^2$, which quantifies the squared difference between values of neighbouring nodes, weighted by the corresponding edges. This term can be interpreted as a measure of the "smoothness" of the node values.

To interpolate missing node values under the assumption of a smooth transition across

the graph, the following minimisation problem is formulated. Suppose there is a known subset of node values x_K and an unknown subset x_U. These can be combined as:

$$x = \begin{pmatrix} x_K \\ x_U \end{pmatrix} \tag{7.5}$$

Similarly, the Laplacian matrix can be partitioned as:

$$L = \begin{pmatrix} L_{KK} & L_{KU} \\ L_{UK} & L_{UU} \end{pmatrix} \tag{7.6}$$

where subscripts indicate the type of node pair considered, e.g., KK denotes edges connecting known nodes. The formulation seeks to find the values of x_U that satisfy the following minimisation problem:

$$\underset{x_U}{\text{minimise}} \quad x^T L x \tag{7.7}$$

This minimisation problem has a closed-form solution:

$$x_U = -{L_{UU}}^{-1} L_{UK} x_K \tag{7.8}$$

Once x_U is computed, the full vector of node values $x = \begin{pmatrix} x_K \\ x_U \end{pmatrix}$ obtains the interpolated values using the basic form of the graph Laplacian method. Although multiple variations of this technique exist, as with the IDW method, this chapter focuses on the fundamental formulation. The aim is to examine how the intrinsic assumptions of these standard methods align with the process of counting transient cycles and reflections within the network.

7.2.9 Leave-one-out cross-validation

Given the limited number of data points available for interpolation, it was essential to utilise as many observations as possible for fitting the methods. To evaluate the generalisation error of the various interpolation approaches, this chapter implements a leave-one-out cross-validation (LOOCV) strategy. This technique has been shown to yield reliable accuracy estimates when applied to datasets with short autocorrelation ranges (Milà et al., 2022). Considering that the present dataset exhibits negligible autocorrelation, LOOCV was deemed an appropriate choice for model validation.

Leave-one-out cross-validation consists of partitioning a dataset $D = \{x_1, x_2, ..., x_n\}$ with n samples into n training subsets $D_{train,i} = D \setminus \{x_i\}$, such that each subset excludes a single sample x_i. This excluded sample forms the test set $D_{test,i} = x_i$. Iteratively, each training subset $D_{train,i}$ is employed to fit a model h_i, which is subsequently evaluated on its corresponding test sample x_i to compute the error metric $e_i = Error(h_i(x_i), y_i)$, where y_i denotes the true value for x_i. After performing all n iterations, the average error metric, defined as $LOOCV, Error = \frac{1}{n}\sum_{i=1}^{n} e_i$, is computed using the results of all subsets. In this chapter, the mean absolute error (MAE) and mean absolute percentage error (MAPE) are utilised as performance metrics for assessing the accuracy of the investigated interpolation methods.

7.2.10 Transient model calibration

This chapter proposes to utilise transient pressure simulation results to guide the estimation of DP_{high} for pressure transients originating from a large industrial user. The underlying concept is to identify a threshold for DP_{high} that yields a strong linear correlation between DP values derived from the transient model (DP_{model}) and those measured above the determined threshold (DP_{high}). This linear relationship is based on the notion that DP_{high} measured in the field constitutes a scaled representation of DP_{model}. While DP_{model} is obtained from simulating a single transient event from inception to dissipation,

DP_{high} encompasses all transient events over a predefined time window. The key requirement of this approach is that the transient model accurately simulates the propagation and reflection of pressure waves originating from the source.

To ensure that the transient model reliably captures the wave dynamics, it was calibrated using a single, randomly selected transient event from the measured data. The calibration procedure was as follows: (1) a genetic algorithm was employed to determine parameters that best matched the rising edge of the initial transient wave in the model with that of the measured data recorded at the sensor with the largest transient magnitude; (2) then, the calibrated model's waveforms were qualitatively and visually compared against the measured waveforms at all sensor locations. In this case study, this comparison showed that while the model reproduced the general wave shape at every sensor, the simulated pressure waves dissipated and dispersed more slowly than observed in the field data.

This discrepancy can be explained by the inherently complex and multifaceted nature of wave dissipation mechanisms, which current state-of-the-art transient solvers do not fully represent (Riaño-Briceño et al., 2022, 2021). As a pragmatic solution, and noting that most pipes in the study area are composed of a single material, a simple post-processing step was applied to the transient model outputs. Specifically, two universal stretching and scaling factors were employed to adjust wave dissipation and reflections. Although this approach is not generally recommended, it produced satisfactory results for the scope of the current investigation. Further details concerning the transient model calibration procedure are provided in Appendix A.

7.3 Case study

This section describes the case study utilised in this chapter. The selected network is a district metered area (DMA) composed predominantly of cast iron pipes, with a mean diameter of approximately 150 mm (Figure 7.2). The network contains an industrial user that generates large, distinguishable pressure transient events ranging from 10 to 20 m

(Figure 7.2). The data were obtained from two monitoring campaigns. The first campaign took place around February 2012 and lasted for two weeks, employing 7 high-speed pressure loggers. The second campaign occurred in June 2012 for a similar duration, using 10 sensors, with 4 of these deployed at the same locations as during the first monitoring period. Figure 7.2 provides further details on the network configuration and sensor placements.

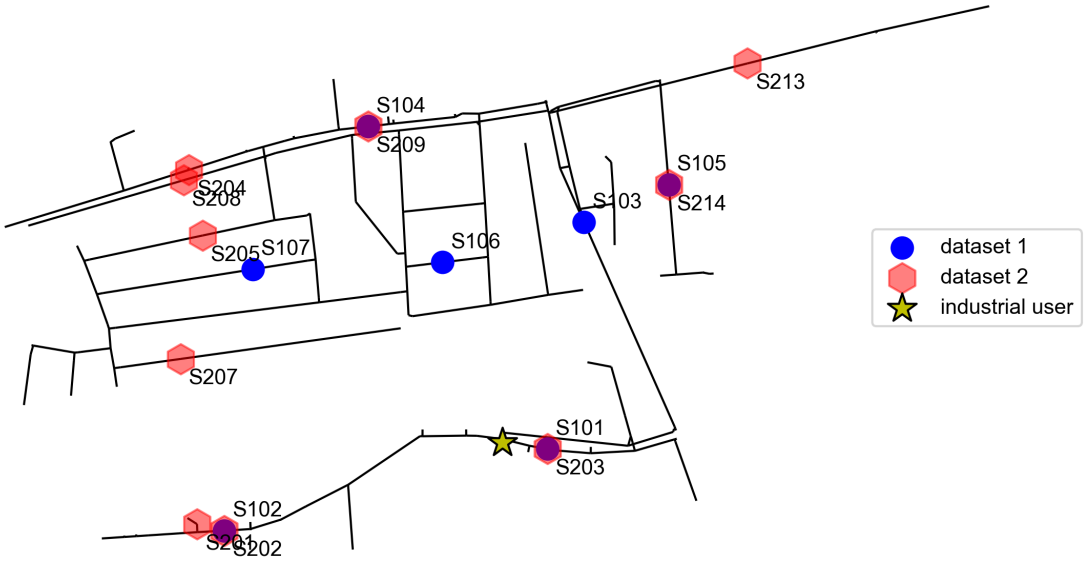
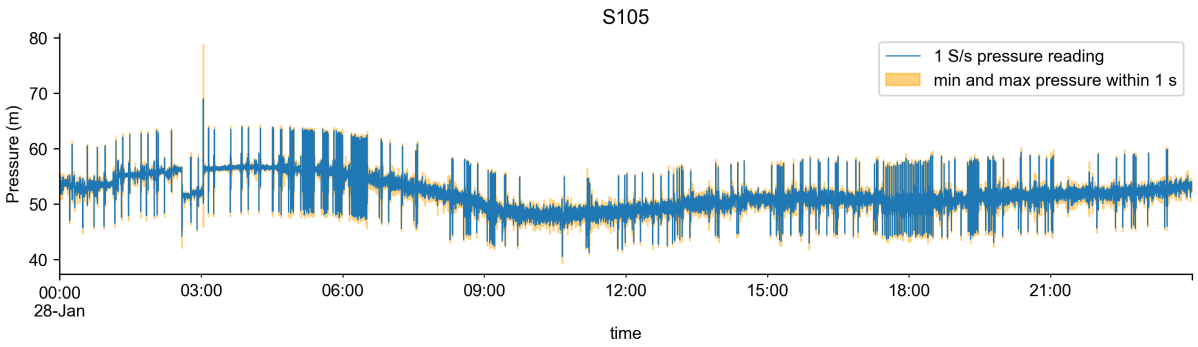


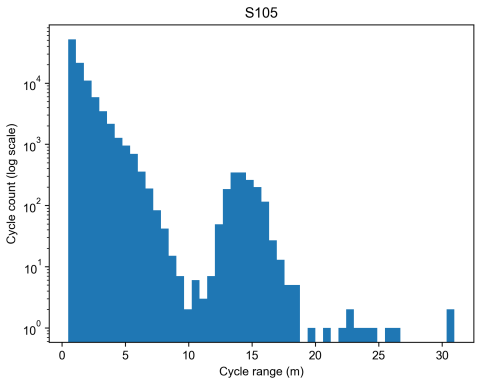
Figure 7.2: Case study network and sensor's location.

7.3.1 Transient pressure profiles

An illustrative daily pressure profile recorded at high temporal resolution is shown in Figure 7.3. The original sensor data, acquired at 64 samples per second (64 S/s), was downsampled to 1 sample per second (1 S/s) for visualisation, while the original maximum and minimum pressure values within each second are indicated separately. Chapter 6 examined the effects of downsampling on cycle counting accuracy using two one-year datasets acquired near a pressure-reducing valve. The findings revealed that reducing the sampling rate from 128 S/s to 1 S/s resulted in slightly over 10% loss of cycles



(a)



(b)

Figure 7.3: Example high-resolution pressure data downsampled at 1 sample per second. Sensor S105. a) 1 day time series. b) one day cycle count histogram.

greater than 10 m, and a much larger discrepancy exceeding 50% for cycles ranging from 2 m to 3 m. Similarly, in this case study, the comparison in cycle loss associated with downsampling from 64 S/s to 1 S/s showed substantial differences in the smaller cycle range. Consequently, in this chapter, the 64 S/s data was utilised to minimise error in detecting and quantifying these smaller cycles.

Figure 7.3a illustrates the pronounced pressure cycles generated by the industrial user, which are easily distinguishable from other, less pronounced hydraulic fluctuations observed in the time series. Additionally, smaller transient events are apparent and are assumed to originate from small commercial users or other diverse customer connections. Notably, the magnitude of these smaller transients declines from midnight to 6:00 AM, a pattern that can be attributed to reduced water use during these hours and hence fewer hydraulic fluctuations. This observation is consistent with the data presented by Marsili et al. (2021).

By calculating and visualising the frequency of pressure cycles within the time series (Figure 7.3b), two distinct cycle distributions emerge, indicating that these cycles may stem from different generating processes. The objective here is to identify a threshold that segregates the pressure cycles into "high" and "low" categories. This distinction enables correlation of a DP_{high} metric, derived from the larger cycles, with a DP_{model} metric obtained from simulating transient events using a transient model. If a strong correlation between DP_{high} and DP_{model} is established, then the transient model outcomes can be utilised to determine DP_{high} and to compare results with spatial interpolation methods for DP_{high}. Meanwhile, DP_{low} would be estimated solely via spatial interpolation.

As an initial step to explore the appropriate threshold for distinguishing DP_{high} and DP_{low}, Figure 7.4 examines how DP values change at each sensor in Dataset 1 when progressively excluding pressure cycles below certain magnitudes. Across all sensors, DP remains relatively stable once pressure cycles below approximately 6 m are excluded, suggesting that only large, industrially generated cycles remain. Considering pressure cycles below 4 m, DP begins to increase gradually for most sensors, and for cycles below 2 m, the magnitude of DP increases abruptly. The threshold for defining DP_{high} and DP_{low} should be as low as possible while still allowing DP_{model} to correlate well with DP_{high}. In this specific case study, the threshold is between 2 m and 6 m.

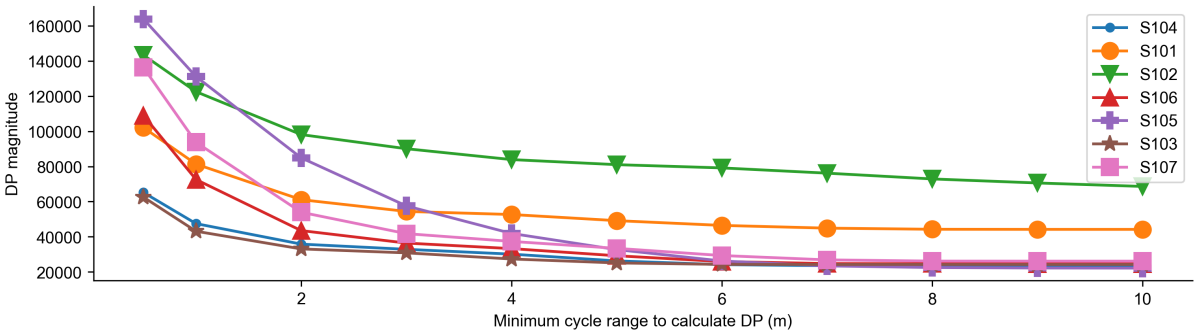


Figure 7.4: DP magnitude variation with increasing minimum cycle range. Dataset 1.

7.4 Results

This section presents the main findings of this chapter regarding the estimation of the dynamic pressure (DP) metric in a water distribution network, based on a given set of sensor measurements.

7.4.1 Determination of the DP separation threshold

This chapter explores the feasibility of decomposing the DP variable into two components: one composed of large, repetitive cycles and another representing smaller pressure fluctuations. The goal is to enhance the interpolation accuracy of the DP variable by identifying an appropriate threshold that distinguishes between large and small cycles. Two distinct datasets were employed for this analysis, and the threshold determination was carried out independently for each dataset.

The analysis of both datasets (Figure 7.5) indicates that a 5 m threshold effectively linearly correlates DP_{model} obtained from the calibrated transient model and DP_{high} in both datasets, despite the model's calibration being limited to a single transient event. This finding is reasonably consistent with the cycle distribution depicted in Figure 7.3b, where a marked decrease in the frequency of cycles is observed above approximately 7 m. The slightly lower 5 m threshold could be explained by the presence of transient events within the 5–7 m range that largely originate from the industrial user, whereas smaller cycles (below 5 m) comprise a mixture of reflections from the large industrial user and other transient sources of varied nature.

Figure 7.6 illustrates the partitioning of DP into components defined by cycles above and below 5 m, applying a minimum threshold of 0.5 m for both datasets. The time periods considered for the DP calculations in the two datasets were identical and included the same distribution of weekdays (Monday to Sunday). Nevertheless, a considerable increase in total DP was observed in Dataset 2 compared to Dataset 1. This increase is consistent with the industrial user's influence, as higher operational frequencies result in elevated

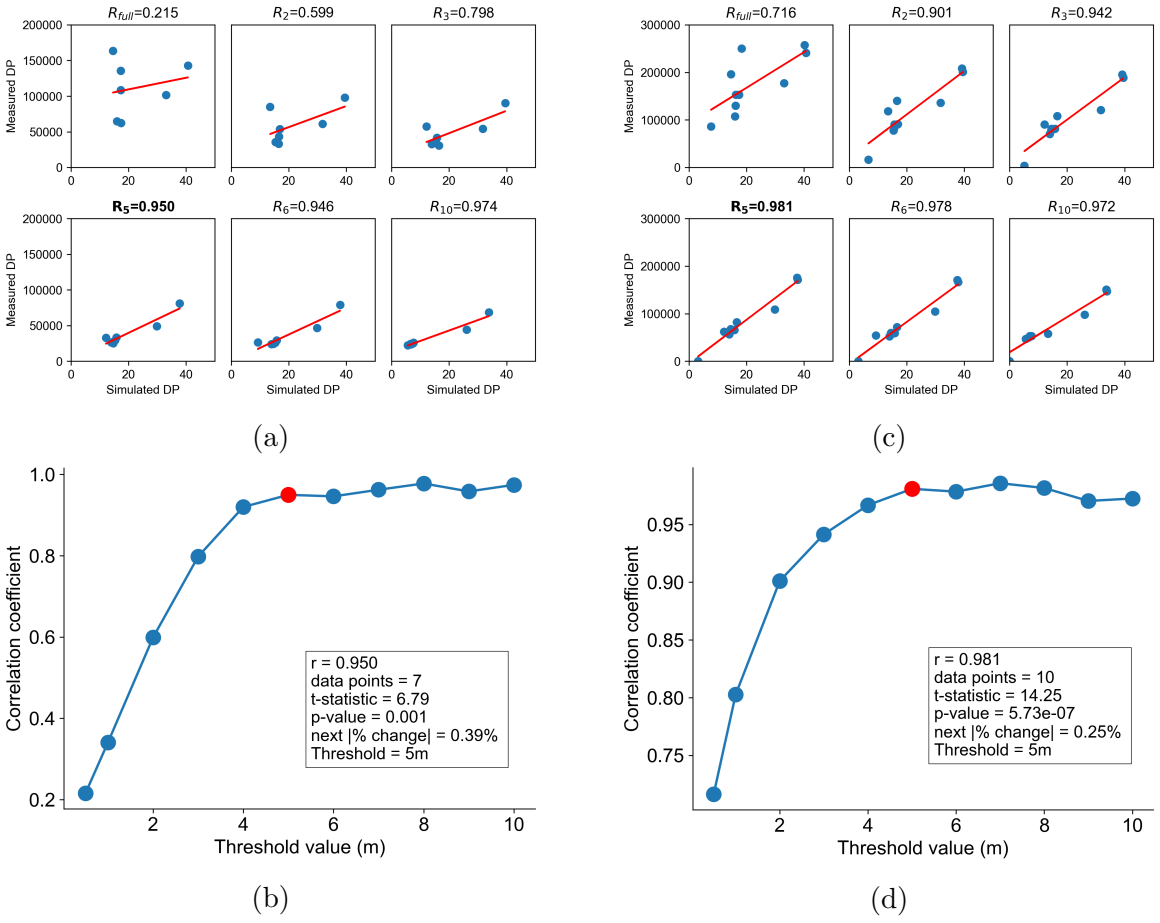
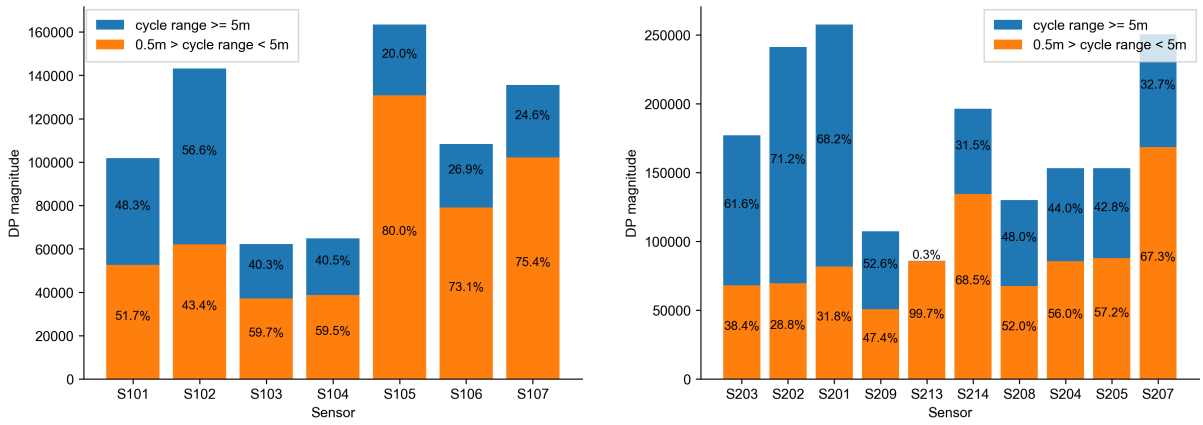


Figure 7.5: Estimation of DP separation threshold. a)-b) Dataset 1, c)-d) Dataset 2.

DP_{high} values. Specifically, for sensor locations common to both monitoring periods, the following percentage differences were found for DP_{high}; 75.6% (S203 and S101), 71.7% (S202 and S102), 73.1% (S209 and S104), and 61.65% (S214 and S105), suggesting a relatively consistent scaling factor across these sensor nodes.

In contrast, DP_{low} exhibits less pronounced discrepancies between the two monitoring periods. For the same common sensor locations, the percentage differences in DP_{low} were 25.6% (S203 and S101), 11.3% (S202 and S102), 27.2% (S209 and S104), and a mere 2.8% (S214 and S105). These smaller differences support the hypothesis that isolating larger transients reveals a relatively stable background level of transient pressure. To further investigate this behaviour within the case study, Figure 7.7 examines the daily variations of both DP_{high} and DP_{low}.



(a) (b)

Figure 7.6: DP separation after 4m threshold setting. a) Dataset 1. b) Dataset 2.

Figure 7.7 reveals considerable day-to-day variability in DP_{high} for both Datasets 1 and 2, whereas DP_{low} demonstrates a more consistent and stable pattern. These findings support the notion that a multitude of users contribute to a persistent background level of transient activity within the network, maintaining relatively uniform magnitudes over time. In contrast, larger pressure cycles appear more variable, likely influenced by the operational patterns of major consumers and the operation of hydraulic control valves.

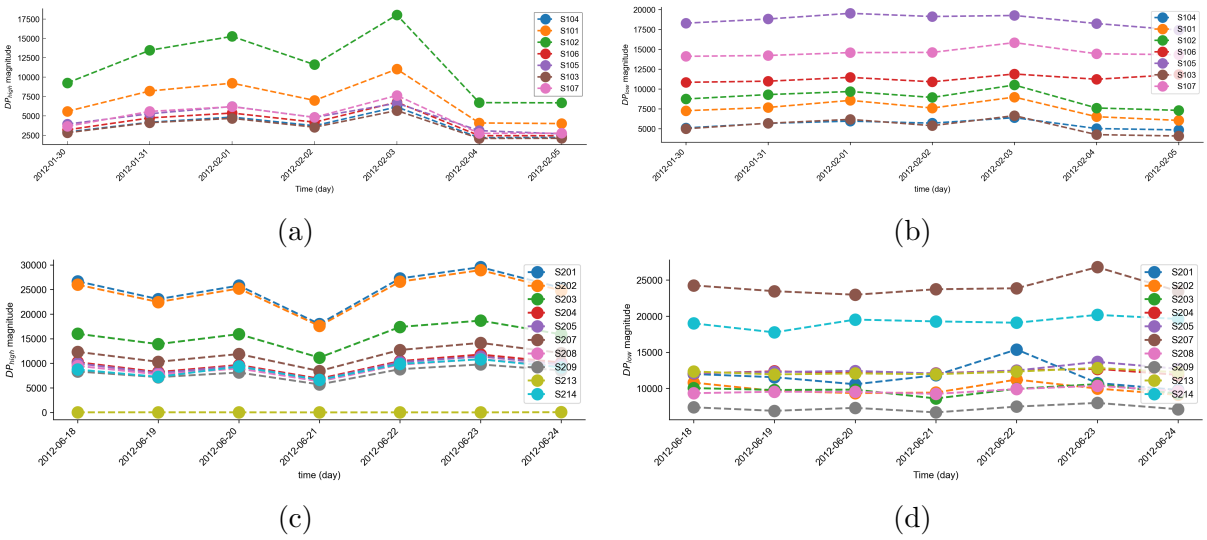


Figure 7.7: Daily DP variations. a) DP_{high} Dataset 1. b) DP_{low} Dataset 1. c) DP_{high} Dataset 2. d) DP_{low} Dataset 2.

7.4.2 Variograms for DP variables

This section presents the results and support the discussion provided in Section 7.2.6 for the variogram analysis used to determine the best suitable interpolation methods for the DP variables after the threshold determination.

7.4.3 Interpolation results

The determined threshold was applied to partition the DP variable into DP_{high} and DP_{low} components, and subsequently each component was interpolated independently. Based on the initial hypothesis of this chapter, DP_{high} can be more accurately estimated using outputs from a transient model, while DP_{low} relies solely on spatial interpolation methods. To evaluate the effectiveness of this approach, it was also compared against interpolating DP without any separation, denoted as DP_{full}. Table 7.1 presents the performance of the various interpolation methods for the three metrics (DP_{high}, DP_{low}, and DP_{full}). Figure 7.9 and Figure 7.10 illustrate the results of the interpolation methods. Each unmeasured node is coloured according to its interpolated DP value. Nodes with measured DP are highlighted with their labels. Additionally, the estimated DP value for all pipes is illustrated as the average of the DP values of their corresponding nodes.

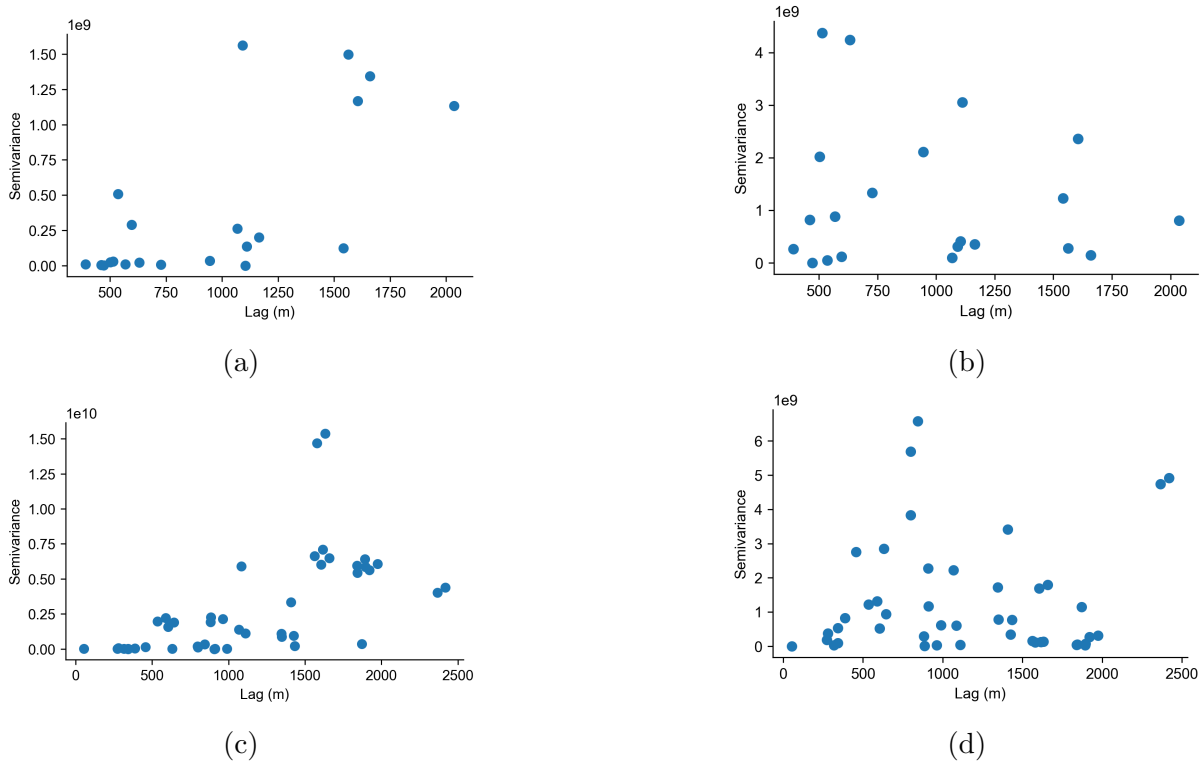


Figure 7.8: Variograms for DP variables. a) DP_{high} dataset 1. b) DP_{low} dataset 1. c) DP_{high} dataset 2. d) DP_{low} dataset 2.

Table 7.1: Leave-one-out cross-validation evaluation for interpolation methods.

Metric	Method	Dataset1		Dataset 2		Dataset 2* (∼S213)	
		MAE	MAPE	MAE	MAPE	MAE	MAPE
DP_{high}	Scaled transient model	**8,463**	**18.90**	**9,786**	**552**	10,151	**10.81**
	IDW interpolation	9,047	19.38	16,503	3,153	**9,621**	12.07
	GL-based interpolation	10,148	23.53	14,906	2,646	8,809	11.82
DP_{low}	IDW interpolation	**30,636**	**48.55**	**25,314**	**26.06**	**27,781**	**28.52**
	GL-based interpolation	33,974	55.01	29,427	31.58	32,885	35.14
DP_{full}	IDW interpolation	39,039	39.07	38,772	26.89	**33,365**	**18.66**
	GL-based interpolation	43,813	44.29	39,778	27.33	34,860	20.32
	IDW + transient model	**38,587**	**38.19**	**33,636**	**19.34**	35,561	19.51
	GL + transient model	39,837	39.85	38,175	22.74	41,139	23.32

DP_{high} interpolation results

The interpolation results of DP_{high} for Datasets 1 and 2 are presented in Figure 7.9. In both datasets, DP_{high} was estimated using the regression line derived earlier to represent the linear correlation between DP_{model} and DP_{high} at the selected threshold. Additionally, the performance of this approach was compared against Inverse Distance Weighting (IDW) and Graph Laplacian-based (GL) interpolation methods.

For Dataset 1, employing the transient model to guide DP_{high} estimation provided slightly improved results compared to the spatial interpolation methods. Specifically, a 6.5% reduction in Leave-One-Out Cross-Validation (LOOCV) Mean Absolute Error (MAE) was observed when comparing the transient model approach to IDW, with an accompanying marginal reduction of 0.48 percentage points in LOOCV Mean Absolute Percentage Error (MAPE).

Initial results for Dataset 2 indicated that scaling the transient model outputs substantially outperformed the other methods in terms of LOOCV MAE. However, the LOOCV MAPE values were generally higher for all three methods. A closer examination revealed that Sensor 213 did not record any large transients exceeding 5 m, potentially due to erroneous readings or specific local hydraulic conditions (e.g., a particular fitting or pipe material configuration) that heavily attenuated the pressure wave. Removing this sensor from the analysis (Dataset 2*) yielded more balanced results across the three methods. Although the scaled transient model still offered a marginal MAPE advantage of about 1 percentage point over the GL method, its performance in terms of MAE was slightly lower. Compared to Dataset 1, Dataset 2* exhibited smaller MAPE values, which is expected due to the increased number of sensors and, consequently, more comprehensive information.

In terms of interpolation capability, the transient model-based approach produced a broader range of DP_{high} values, extending beyond the domain of observed measurements (Figure 7.9). While extrapolation beyond observed data is generally not recommended, it

is justified here based on the initial assumption that DP_{high} should exhibit a strictly linear relationship with DP_{model}. By contrast, the spatial interpolation methods are inherently limited to predicting values within the range of observed data.

In all cases, the highest DP_{high} values were localised in close proximity to the transient source. The distribution of elevated DP_{high} magnitudes was influenced by the transient source and a series of dead-end pipes in the lower left region of the network, which generated multiple reflections. Such reflections significantly affect DP values, underscoring the importance of accurately defining boundary conditions in the transient model to achieve reliable results using the transient model-based approach.

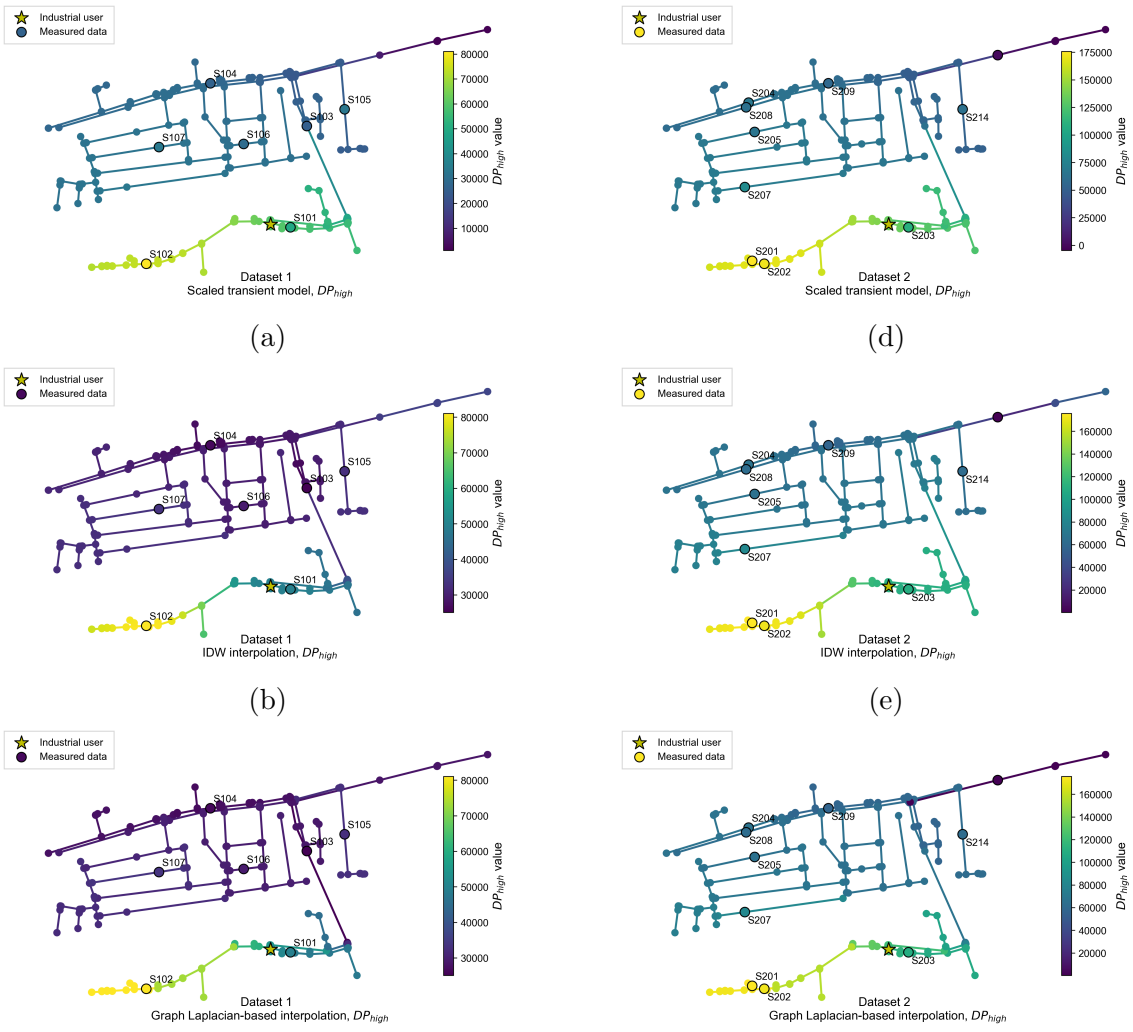


Figure 7.9: DP_{high} interpolation. a)-c) Dataset 1, d)-f) Dataset 2.

DP_{low} interpolation results

DP_{low} is intended to capture the influence of numerous smaller transient events originating from various sources on the spatial interpolation methods. Since no suitable transient model is available to simulate this scenario, DP_{low} must be estimated exclusively through spatial interpolation methods. Based on a review of existing literature, IDW and a Graph Laplacian-based interpolation methods were selected to obtain DP_{low} values at unmonitored locations. Figure 7.10 presents the corresponding interpolation results.

The results indicate that, for this case study and both datasets, higher DP_{low} values are mainly concentrated in the left and right portions of the network. However, because Dataset 2 includes a greater number of monitoring locations than Dataset 1, it provides a more detailed representation of small magnitude, recurrent pressure cycles. In particular, Sensor 207 in Dataset 2 captured the highest DP_{low} value network-wide, a detail that was undetected in Dataset 1 due to the absence of a sensor at that location. Overall, IDW yielded better performance than the Graph Laplacian-based interpolation method across both datasets. Nonetheless, the MAPE values were substantially higher for DP_{low} estimation compared to DP_{high}. For Dataset 1, the MAPE reached 48.55% using IDW, which decreased to 28.52% in Dataset 2, likely influenced by the larger number of sensors and the improved spatial coverage they provide.

DP_{full} interpolation results

In addition to examining DP_{high} and DP_{low} separately, the direct interpolation of DP (DP_{full}) was also evaluated employing the established lower limit of 0.5 m for the pressure cycle range. The results of this assessment indicate that interpolating DP_{full} effectively balances the individual outcomes obtained when DP_{high} and DP_{low} are considered independently. Overall, the spatial interpolation methods demonstrated a performance comparable to the transient model-based approach for estimating DP_{high}, yielding nearly equivalent accuracy levels.

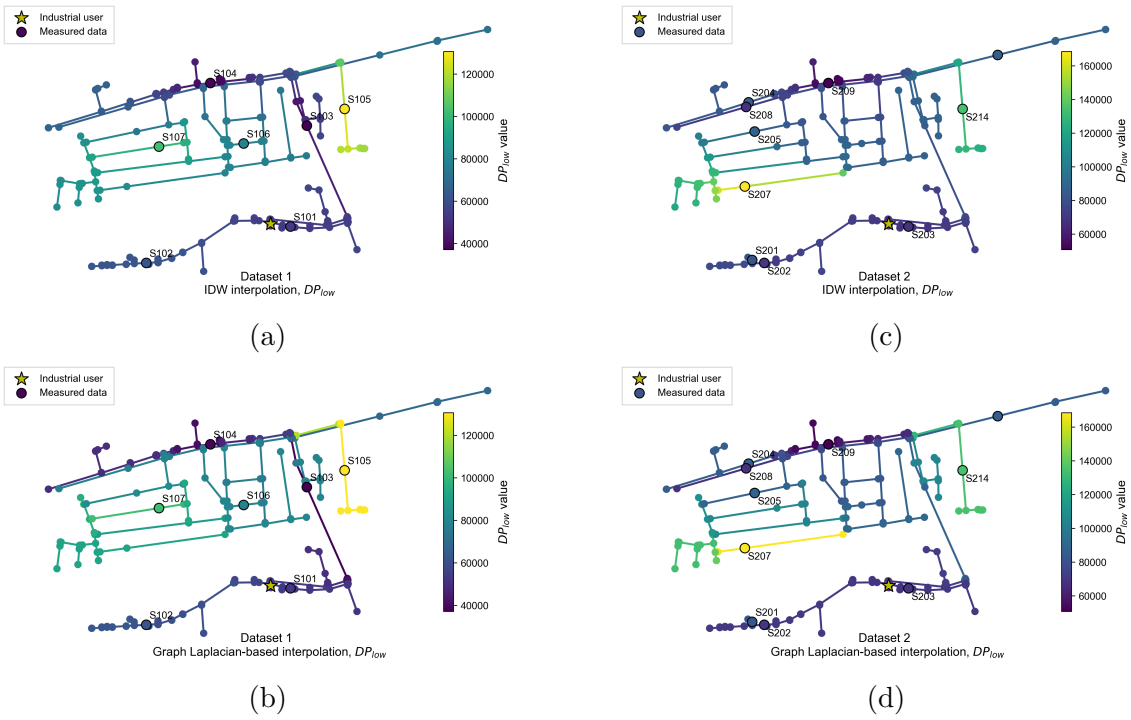


Figure 7.10: DP_{low} interpolation. a), b) Dataset 1; c), d) Dataset 2.

7.5 Discussion

A comparison of the scaled transient model with the spatial interpolation methods reveals that, although the scaled transient model achieves the lowest mean absolute percentage error (MAPE), the improvement is marginal. This finding suggests that spatial interpolation can provide reliable DP_{high} estimates if an appropriate separation threshold is determined. However, it is important to note that the scaled transient model approach can produce a broader range of values, including those outside the observed domain, which is an advantage that spatial interpolation methods cannot offer, especially when only a limited number of sensors is available. Additionally, a transient model linked to the source of large transients remains necessary to identify a suitable threshold based on DP_{model}.

If a utility lacks the capacity to simulate transient pressures, an approximate threshold could be established through probabilistic analyses of large transient events. To illustrate, the sensitivity of DP_{high} contribution to total DP was tested by changing the threshold from 5 m to 10 m, observing up to a 30% difference in Dataset 2. This indicates that

isolating large repetitive transients from smaller, more diverse instabilities can still yield reasonably accurate DP_{high} estimates, even if the threshold is less precisely determined. The trade-off is that this approach, however, increases DP_{low} and its associated interpolation error.

Because DP_{low} reflects the influence of numerous smaller transient sources, transient models cannot be employed to estimate DP_{low} at unmonitored locations, and reliance on spatial interpolation methods is inevitable. This analysis suggests that the accuracy of DP_{low} estimation is more sensitive to the number of monitoring sensors than DP_{high}. Results from Dataset 2* show that even a modest increase in sensor coverage notably reduces MAPE. This raises questions about the minimum sensor density or spacing required for reliable spatial interpolation of DP. Furthermore, the incorporation of additional network connectivity metrics (Marsili et al., 2023) could improve the interpolation of both DP_{high} and DP_{low}.

In networks affected by large repetitive transients from a small number of sources, separating their contribution using a threshold provides a more reliable estimation of DP. If only transients above a certain magnitude comparable to the determined threshold are relevant for analysing pressure-induced pipe fatigue, interpolation methods such as inverse distance weighting can be effectively applied to estimate DP at locations lacking direct measurements. Such approaches enable the development of dynamic pressure maps, which, when combined with mean pressure measurements, enhance our understanding of cumulative pressure-induced stresses on pipes. This information can be integrated into asset management tools that estimate the probability of pipe breaks.

This chapter defined dynamic pressure as the product of pressure cycle magnitudes and their repetition frequency, following the promising results presented in Chapter 7. However, as presented in the Literature Review (Chapter 2) improvements may be possible by testing exponential models of the form $\sum_{i=1}^{M}(\sigma_i \cdot n_i^m)$ using pipe failure data. Such models would place greater emphasis on larger pressure cycles, potentially providing a more realistic representation of fatigue effects consistent with S-N curve principles.

7.6 Conclusion

This chapter examined the feasibility of interpolating the dynamic pressure variable throughout a district metered area. The dynamic pressure metric, defined as the product of pressure cycle frequency and range, captures transient activity within water distribution networks. The characteristics of the analysed datasets enabled a decomposition of the dynamic pressure variable into two components: one associated with large pressure cycles, and another corresponding to smaller fluctuations. In the presented case study, the larger pressure cycles originated from a single source, facilitating their representation using a transient pressure simulator. This premise allowed for a framework in which interpolation methods were evaluated separately for DP_{high}, influenced by a single source, and DP_{low}, representing more diverse transient sources.

For DP_{high}, the interpolation involved scaling the outputs of a transient model and comparing these results to conventional interpolation techniques. In contrast, DP_{low} was estimated solely using spatial interpolation methods. The principal findings indicate that spatial interpolation techniques, particularly Inverse Distance Weighting, performed comparably to the transient model approach for estimating DP_{high}. Although the interpolation of DP_{low} exhibited higher errors in Dataset 1, these were substantially reduced in Dataset 2 by adding just two more sensors. Overall, this chapter demonstrates that spatial interpolation can reliably estimate dynamic pressure at unmonitored locations, provided that a sufficient number of sensors capture the key pressure dynamics of the network. Future research should aim to identify the optimal sensor density and spatial arrangement to ensure consistent accuracy across various water distribution systems.

The methodologies introduced in this chapter enable water distribution network operators to continuously characterise the dynamic hydraulic conditions within their systems using data from high-speed pressure monitoring devices. By integrating these measurements to estimate the Cumulative Pressure-Induced Stress (CPIS) for each pipe, utilities can incorporate these dynamic stress indicators into pipe break prediction and causality models. Such analysis allows for the precise identification of the most vulnerable segments in the

network, supporting proactive maintenance and rehabilitation efforts.

The ability to quantify CPIS and understand its relationship with pipe deterioration rates provides a robust scientific basis for implementing targeted hydraulic calming and pressure control measures. These interventions, supported by reliable, high-frequency monitoring data and advanced modelling techniques, not only can extend the service life of pipe infrastructure but also would enhance overall network resilience and reliability. In turn, this approach helps reduce pipe breaks, water losses, and repair costs, ultimately contributing to more sustainable and efficient management of water distribution networks.

Chapter 8

Localisation of sources of pressure transients

Related publication: Jara-Arriagada, C, Ulusoy, A.J., & Stoianov, I. (2024) Localization of Transient Pressure Sources in Water Supply Networks with Connectivity Uncertainty. *Journal of Water Resources Planning and Management.* 150(4), 107731. doi:10.1061/JWRMD5.WRENG-6235

Submitted manuscript: Jara-Arriagada, C, & Stoianov, I. (2025) Field validation of a pressure transient source localization method in a water distribution network. Under Review. *Journal of Hydroinformatics.* ISSN: 1464-7141

8.1 Introduction

Chapter 7 presented a framework for the interpolation of CPIS (on the basis of the DP metric) in water distribution networks. This framework requires identifying sources of large pressure transients. As presented in the literature review (Chapter 2) various approaches for pressure transient source localisation has been developed. Most of these approaches utilise differences of pressure wave arrival times to pinpoint the source of a transient. However, a critical drawback of current approaches is that these do not account

for uncertainty in network connectivity. This uncertainty refers to the potential mismatch between the network connectivity of a hydraulic model and the actual connectivity in the operational network. Such mismatch can often occur due to improper records of boundary valves, blockages in network, or large errors in assumed wave speed in materials. As the approaches for transient source localisation use pressure wave speeds, small errors in the network connectivity can produce large errors in the results. As a solution to this problem, this chapter proposes a novel method for locating sources of pressure transients considering uncertainty in network connectivity.

The method proposed in this chapter formulates a mixed-integer quadratic problem that considers the k-shortest (fastest) paths between pressure transient sources and different pressure sensors. This formulation results in a large combinatorial problem, which is solved to select the most likely source of a pressure transient. The presented formulation of the pressure transient source localisation problem not only provides a solution for identifying the most likely source of a transient, but it also offers additional benefits for validating network connectivity and detecting unreported blockages.

The performance of the proposed method is assessed in a first instance through a comprehensive set of simulated scenarios involving pressure transient events. Then, an initial investigation on validating the method is performed using transient records from a real case study network. The results of this chapter demonstrate the improved robustness and the importance of considering multiple k-fastest paths for pressure wave propagation in cases with uncertain network connectivity. The remainder of this chapter describes the problem formulation, followed by the solution method, and the experimental setup to solve the formulated problem. Subsequently, details of the case studies and the outcome of this research are presented. The chapter concludes with insights derived from this research and suggestions for further validation and implementation of the presented method in operational networks.

8.2 Methodology

This section details the methodology developed to address the challenge of localising the source of pressure transients in networks with uncertain connectivity. First, the evolution of the problem's formulation is introduced, transitioning from a timing-based scoring algorithm to an optimisation problem. Following that, the strategy used to solve the optimisation problem, along with the accompanying methods are presented.

8.2.1 Problem formulation

This section describes the proposed formulation for the localisation of the source of pressure transients in water distribution networks with uncertain connectivity as an optimisation problem. First, the basic concept of a timing-based scoring algorithm is presented as documented in the literature. The scoring algorithm is then reformulated as an optimisation problem. Next, the optimisation problem is extended to consider multiple propagation paths for the first pressure wave from the the source to different sensors locations. The solution to this optimisation problem encompasses both the source of the pressure transients and the propagation paths taken by the first wave to reach installed pressure sensors. Specifically, the solution identifies the best match between the time differentials of estimated wave arrival times along chosen routes to the sensor positions and the actual wave arrival times recorded at the sensors. Using this approach, uncertainties in network connectivity are addressed. Finally, a weight function is applied to improve the performance of the localisation algorithm.

Timing-based scoring method for pressure transient source localisation

The scoring approach for timing-based pressure transient source localisation relies on the definition of a score S for every potential transient source s within the set of nodes N in a WDN as (Srirangarajan et al., 2010; Zecchin et al., 2022):

$$S_s = \sum_{i,j \in M} \left| (t_i - t_j) - (T_{is} - T_{js}) \right|, \forall s \in N \tag{8.1}$$

where M is the set of sensors nodes that detected the transient the event, t_i is the measured time when the first wave of the transient event is detected by the sensor i, and T_{is} is the analytical time taken by the pressure wave to travel between node i and the potential source. Namely, the distance is divided by the wave speed (celerity). The above scoring function (Equation (8.1) reads as follows: for all pairs of sensors i and j that belong to the set of sensors M, the score of a node s will be the difference between the measured pressure wave arrival time to the sensor i and a sensor j minus the difference of the analytical time of arrivals from a source s to the sensors i and j (Srirangarajan et al., 2010; Srirangarajan & Pesch, 2013). The node s which provides the minimum score is then selected as the pressure transient source (Srirangarajan et al., 2010; Srirangarajan & Pesch, 2013). The analytical time of propagation of a transient wave from a source s to a sensor i, can be estimated as follows:

$$T_{is} = \frac{d_{is}}{u_{is}} \tag{8.2}$$

where d_{is} is the distance between the source s and the node i and u_{is} is the average wave speed between the source s and node i. For the sake of simplicity in this study, it is assumed that all pipes have a known and equal wave speed. It should be noted that when applying the discussed method to operational networks, one must account for uncertainties in wave speed and ensure precise determination of the initial pressure wave's arrival times at the sensors. The challenges and constraints of the numerical experiments presented here are further elaborated upon in the Conclusion section.

In order to calculate the analytical times of wave arrivals, graph theory is used to find the fastest path between the potential source s and all the sensors. The Dijkstra's algorithm (Dijkstra, 1959) is commonly applied for finding the shortest path between two points in a network. In order to determine the fastest paths, the length of the pipes are divided

by the wave speed corresponding to the pipe material, and the Dijkstra's algorithm is applied, using these weighted pipe lengths, to select the paths with shortest travel time.

Reformulation of the timing-based pressure transient source localisation in networks as an optimisation problem

The timing based scoring approach can be reformulated as a mixed-integer quadratic program, where the objective is to minimise, over all potential sources s, the error between the difference of measured times of pressure wave arrivals at i, j against the difference of analytical wave arrival times between the source s and sensors i, j. The next section shows that the optimisation problem can be efficiently expanded to represent more complicated connectivity scenarios (e.g. pipe blockages).

The formulation of the scoring approach as an optimisation problem is as follows:

$$\underset{z}{\text{minimize}} \quad \sum_{i,j \in M} \left[\sum_{s \in N} (T_{is} - T_{js})\, z_s - (t_i - t_j) \right]^2 \tag{8.3a}$$

$$\text{subject to} \quad \sum_{s \in N} z_s = 1, \tag{8.3b}$$

$$\sum_{s \in N} (T_{is} - T_{js})\, z_s \geq 0, \forall i, j \in M \text{ s.t. } t_i \geq t_j, \tag{8.3c}$$

$$z_s \in \{0, 1\}, \forall s \in N \tag{8.3d}$$

where z_s is a binary decision variable, which takes the value of 1 if the source is in node s or 0 otherwise, M is the set of sensors that detected the transient event, and N is the set of potential source locations (network nodes). T is a multidimensional array of size $n(M)$ by $n(N)$ (where $n(M)$ represents the cardinality of set M) that stores the analytical shortest travel time between a source s and every sensor as defined in Equation (8.2) and t is an array that stores the measured time of detection of the pressure wave by every sensor.

Problem (8.3) has a convex quadratic objective function, linear constraints and integer

decision variables. This formulation results in a mixed-integer quadratic program which can be solved to global optimality with off-the-shelf mixed-integer programming solvers. Constraints (8.3b) and (8.3c) indicate, respectively, that there should be only one source among all nodes in the network and that the solution must satisfy the order of the measured times of arrival of the pressure wave at the different sensors (i.e., the order of analytical arrival times from the source to the sensors must match the order of measured arrival times). In other words, if a sensor j is reached before another sensor i according to the measured arrival times t_j and t_i (i.e., $t_i \geq t_j$), the analytical travel time between j and the selected source s (i.e., $z_s = 1$) is required to be smaller than the analytical travel time between s and i (i.e., $T_{is} \geq T_{js}$ or, equivalently, $T_{is} - T_{js} \geq 0$).

Given the integer problem formulation and constraint (8.3b), the solution of (8.3) will identify a unique source node. Nevertheless, it is possible to relax the problem by removing the binary constraint (8.3d), and setting $0 \leq z \leq 1$ to obtain a score for every node, similarly to the original formulation (Equation (8.1)). In addition, we can add the l^2-norm of z multiplied by a regularisation parameter λ to the objective function to encourage sparsity in the solution vector. The l^2-norm encourages every term in the solution vector to be small, while the regularisation parameter adjusts the extent of this reduction. This relaxed formulation with regularisation is presented below.

$$\underset{z}{\text{minimize}} \quad \sum_{i,j \in M} \left[\sum_{s \in N} \left(T_{is} - T_{js} \right) z_s - \left(t_i - t_j \right) \right]^2 + \lambda \sum_{s \in N} z_s^2 \tag{8.4a}$$

$$\text{subject to} \quad \sum_{s \in N} z_s = 1, \tag{8.4b}$$

$$\sum_{s \in N} \left(T_{is} - T_{js} \right) z_s \geq 0, \forall i, j \in M \text{ s.t. } t_i > t_j, \tag{8.4c}$$

$$z_s \in [0, 1], \forall s \in N \tag{8.4d}$$

The problem's relaxation and the inclusion of a regularisation parameter are introduced here to demonstrate that results similar to the scoring algorithm can be achieved by reformulating the problem as an optimisation problem. Nevertheless, in subsequent iterations

of the optimisation problem, both the relaxation of integer variables and regularisation are excluded due to the intricacy involved in deriving significant solutions. This complexity might arise from the possibility that several nodes, not necessarily near the source, might influence the solution. Further exploration of this solution approach is needed using experimental data from operational networks.

Extension of the optimisation problem to account for networks with uncertain connectivity

In both Problem (8.3) and Problem (8.4), the problem formulation assumes that the fastest paths between the source and the pressure sensors are known. However in networks with uncertainty in connectivity, there is a potential for mistakenly identifying the fastest paths. When hydraulic models do not include the most recent data on closed valves or alterations in pipe connectivity, the outcomes derived from the initial framework for transient source localisation might be inaccurate. To address the issue of uncertain connectivity, the problem formulation can be augmented to account for various possible routes between the source and each pressure sensor. In particular, this research considers the k-shortest (fastest) paths (Yen, 1971) between the source and the pressure sensors, where k is a user-defined parameter. However, integrating k-fastest paths into the algorithm results in a large combinatorial challenge, leading to significant computational demands. While exhaustive enumeration can address the combinatorial issue for smaller networks, the size of the search space in medium to large networks grows exponentially with the number of variables. This renders the exhaustive method infeasible, prompting a shift towards an optimisation problem formulation, which explicitly includes the appropriate constraints. The optimisation problem is formulated as follows:

$$\underset{z}{\text{minimize}} \quad \sum_{i,j \in M} \left[\sum_{s \in N} \sum_{p \in P_{is}} T_{isp} z_{isp} - \sum_{s \in N} \sum_{p \in P_{js}} T_{jsp} z_{jsp} - (t_i - t_j) \right]^2 \tag{8.5a}$$

$$\text{subject to} \quad \sum_{s \in N} \sum_{p \in P_{is}} z_{isp} = 1, \forall i \in M, \tag{8.5b}$$

$$\sum_{p \in P_{is}} z_{isp} = \sum_{p \in P_{js}} z_{jsp}, \forall i, j \in M, s \in N, \tag{8.5c}$$

$$\sum_{s \in N} \sum_{p \in P_{is}} T_{isp} z_{isp} - \sum_{s \in N} \sum_{p \in P_{js}} T_{jsp} z_{jsp} \geq 0, \forall i, j \in M \text{ s.t. } t_i > t_j, \tag{8.5d}$$

$$z_{isp} \in \{0, 1\}, \forall i \in M, s \in N, p \in P_{is}, \tag{8.5e}$$

where z_{isp} is a binary decision variable, which takes the value of 1 if there exists a source at node s and the wave propagated through the path p from s to the sensor i, M is the set of sensors that detected the transient event, P_{is} is the set of k-fastest paths between nodes i and s, and N is the set of potential locations for the transient source (network nodes). Here, T is a $n(M)$-by-$n(N)$-by-k array that stores the analytical travel times of wave propagation along the k-fastest paths between potential sources s and the sensors. Then, T_{isp} represents the wave propagation time to a sensor i from a source s in the path p. The array t of measured arrival times is defined as in (8.3). Constraints (8.5b) and (8.5c) ensure that only one node in the network is selected as the source of the pressure transient, and that, from all k potential paths that can be taken for the wave to reach the sensors, only one is selected for every sensor.

Alternatively stated, with Constraint (8.5b) it is enforced that for all sensors i, the sum of z over all the potential sources s and paths p is equal to 1. Therefore, a single source s will be selected for every sensor. To ensure that the source s remains the same for all sensors, Constraint (8.5c) is included. This constraint enforces that the sum of z over all potential paths from a source s to a sensor i has to be the same as the sum of z over all potential paths from the same source s to every other sensor j. Then, if the source node is found in s, the sum of z_{isp} over all paths p between s and i will be 1 and, because of (8.5c),

the sum of z_{jsp} over all paths between s and j must also be 1 for every other node j. Similarly, if the source is not in s, the sum of z_{isp} will be 0, and this will also be the case for the sum of z_{jsp} over all paths p between s and j, for all $j \in M$. Therefore, the same source s will be selected for all sensors. The sum over all paths in both Constraints (8.5b) and (8.5c) also ensure that a single propagation path is chosen from the available paths between potential sources s and sensors i.

The problem formulation described here might face challenges in identifying the precise fastest path if there are inaccuracies in the arrival time of the initial pressure wave or in the model simulating the pressure transient propagation. There are two primary sources of uncertainty in addition to the uncertainty in network connectivity. The first concerns the determination of the pipe wave speeds, which for the purpose of this paper, are presumed known. The second uncertainty arises from pinpointing the pressure wave's arrival times. The sharpness of the pressure change when a pressure wave reaches a sensor has an impact on determining its exact arrival time. A distinct sharp change in pressure allows for straightforward detection of the wave's arrival time. On the other hand, if the pressure wave and transition are dissipated and gradual because of peak demand or other factors, or if the pressure readings are affected by inherent device noise or minor transients reflections within the system, the recorded wave arrival time may exceed the analytical estimation (see Appendix C and Appendix D). These longer measured wave propagation times can cause the solution of the optimisation problem to choose a path that is slightly longer than the actual propagation path taken by the pressure wave in order to compensate the time measuring error. In order to address this problem, a weight function is incorporated into the objective function to discourage the selection of longer k-fastest paths. This weight function increases the analytical time of wave propagation, so that the error in measured time is not compensated by higher order paths in the solution of the optimisation problem. The weight of a path p is given by:

$$w_p = 1 + \alpha\sqrt{l_p} \tag{8.6}$$

where, for a set of paths P, the index set $L(P) = \{0, ..., |P| - 1\}$ is defined such that $l_p \in L(P)$ is the position of path $p \in P$ when the set of paths P is sorted in order of increasing pressure wave propagation times. The value of α scales the increase in time that the weight adds for a specific k-fastest path and needs to be tuned for every network. The weight w_p becomes larger for higher order paths and, by design, increases with the shape of the square root function. Further explanations for the purpose and meaning of the weight function are presented in the Appendix C section of this thesis. The proposed weight is applied to the analytical times of arrivals in the objective function as follows:

$$\underset{z}{\text{minimize}} \quad \sum_{i,j \in M} \left[\sum_{s \in N} \sum_{p \in P_{is}} T_{isp} z_{isp} w_p - \sum_{s \in N} \sum_{p \in P_{js}} T_{jsp} z_{jsp} w_p - (t_i - t_j) \right]^2 \tag{8.7a}$$

$$\text{subject to} \quad \text{same constraints as Problem (8.5)} \tag{8.7b}$$

The optimisation problem described in Problem (8.7) provides a solution for identifying the possible source of transient waves using the differences in arrival times of pressure waves at sensors. Additionally, this formulation also determines the path taken by each pressure wave. By including a weight in the objective function, the solution accounts for possible uncertainties in measured and estimated arrival times. This optimisation problem is a mixed-integer quadratic problem with linear constraints, which can be solved to global optimality using off-the-shelf mixed-integer quadratic programming solvers.

8.2.2 Solution method

This section provides an overview of the procedures that were used for the implementation, solution, and reproducibility of the developed problems described in the problem formulation section. This solution method is also illustrated in Figure 8.1. In particular, the rest of this section focuses on the procedure to solve Problem (8.7), which is analogous for the solution of Problem (8.4) and Problem (8.5).

Initially, network data, sensor locations, pressure time series, and the predefined number of k-fastest paths are inputted. Following this, the wave arrival onset at each sensor is identified using the pressure time series data. Sensors are then sequentially ranked based on their proximity to the source, deduced from the derived pressure wave arrival times. Essentially, the first sensor to record the pressure wave's arrival must be the closest to the source in travel time and is, therefore, positioned at the top rank. This ranking continues in a similar manner for the remaining sensors. The sequence in which sensors are ranked is depicted in the stored arrival times array, denoted as t. Within this array, the quickest arrival time, t_0, corresponds to the top-ranked sensor, and the sequence proceeds likewise for subsequent sensors.

The k-paths of shortest travel time (k-fastest paths) to the sensors for every node in the network are then calculated. This is done by weighting/dividing the length of the pipes by their wave speed, and then applying the k-shortest paths algorithm for every node to the sensors. The results of this procedure are stored in a 3D multidimensional array T in which T_{isp} represents the time taken by the pressure wave to reach sensor i from the source s in the path p. The constant α used in the weight function is defined by trial and error. Using this data, the problem is formulated and solved to global optimality using an off-the-shelf optimisation solver.

The outcome of the optimisation problem identifies the source that best reduces the discrepancy between the observed pressure wave arrival times at sensors and the theoretical arrival times stored in the multidimensional array T. This solution is stored in an array called z of the same shape as T. The value of z_{isp} is 1 if the source is in s and the shortest travel time from s to i is in path p, and 0 otherwise. Based on these results, the location of the node in the network that represents the source of the transient event, as well as the paths the wave travelled from the source to each sensor, can be identified.

If the path order in which a wave travels to any sensor (whether it is the first, second, and so on, fastest path) matches k (the initial maximum number of fastest paths provided to the algorithm) it is inferred that the solution might require considering additional

paths to be accurate. To address this, the number of k-fastest paths is increased by a user-defined amount n, and the procedure is repeated. The algorithm determines the multidimensional array of (k+n)-fastest paths and solves the corresponding optimisation problem. Should the path order for every sensor be less than or equivalent to $k + n$, and if there is no variation in the solution across iterations, it is assumed that the algorithm has successfully identified the source of the pressure transients and derived the paths the waves would have travelled from the source to the sensors.

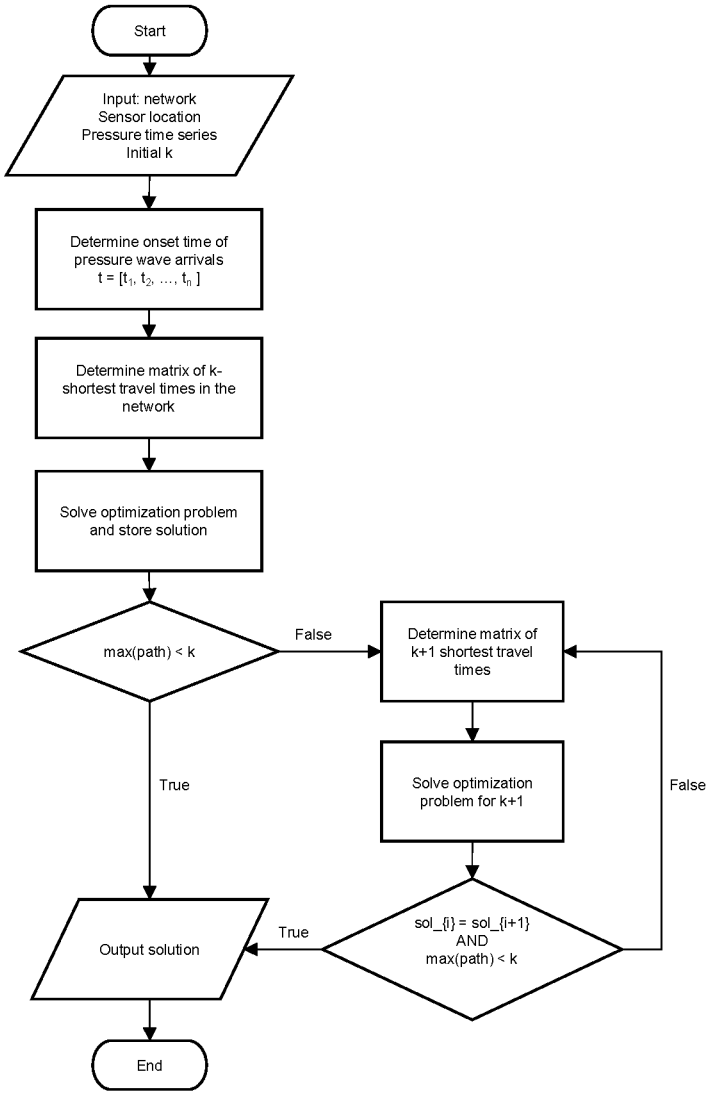


Figure 8.1: Flowchart for new transient source localisation. From: Jara-Arriagada et al. (2024).

8.2.3 Measurement of errors in results

The errors for every path selected by the transient source localisation method have been calculated as the sum of the square error of the difference between the measured times of arrival between sensor j and all other sensors i and the analytical difference between sensor j and all other sensors i:

$$\sum_{i,j \in M} \left[(T_i^* - T_j^*) - (t_i - t_j) \right]^2 \tag{8.8}$$

Where, T_i^* is the analytical pressure wave propagation time from the found source s to the sensor $i \in M$, and t_i is the measured time of arrival of the pressure wave to sensor $i \in M$.

8.2.4 Experimental setup

Pressure transient simulation

The initial investigation of this study has employed simulated data to evaluate the effectiveness of the proposed pressure transient source localisation formulation. This approach facilitates controlled testing of the formulation's performance and an assessment of its error. Pressure transient data were simulated using the TSNet v.0.2.2 software package (Xing & Sela, 2020). This Python-based software implements the method of characteristics to simulate the propagation of pressure transients from a given source. The source of pressure transients can be simulated as pump operations, bursts, large user demands, or valve operations. For the case of pipe bursts, these are simulated at network nodes. The TSNet software was used to generate time series of pressure transients for every node. However, only a subset of nodes M (representing sensors) was utilised to compute pressure transient arrival times t.

Determination of k-fastest paths

The determination of the k-fastest paths in the network from a source s to all nodes i in the network was performed using the function 'shortest_simple_paths' from the package NetworkX v.2.6.3 in Python v.3.7.13. This function uses the procedure to find k-shortest paths presented in (Yen, 1971) and the Dijkstra algorithm (Dijkstra, 1959). In order to get the k-fastest paths, the lengths of the pipes are divided by their wave speed to obtain the travel time. Then, the algorithm for k-shortest paths is applied to determine the k paths of shortest travel time between every pair of potential source $s \in N$ and sensor $i \in M$, and the output is stored in T.

Solution of the Optimisation Problem

The formulated optimisation problems (Problem (8.4), Problem (8.5) and Problem (8.7)) were solved in Python v.3.7.13 using the CVXPY v.1.2.1 package (Diamond & Boyd, 2016; Agrawal et al., 2018) and the commercial solver GUROBI v.9.5.2. (Gurobi Optimization LLC, 2023).

8.3 Numerical experiments

This section describes the experimental framework employed to implement and validate the proposed problem formulation and solution method. The section provides reference to the criteria for detecting the onset of pressure transients, and metrics for error measurement. In addition, all the details needed to ensure the reproducibility of the simulated case studies and the obtained results are presented.

8.3.1 Simulated case studies

Multiple pressure transients originating from different sources and various network configurations were simulated to obtain test data for assessing the efficacy of the developed source localisation method. Three primary scenarios were devised to showcase the methodology, complemented by eight more scenarios for extended validation. All simulated scenarios tested in this chapter are based on the network presented in Figure 8.2, which also shows details of the three main scenarios. This network is a modified version of the Net3 network provided as an example for the software EPANET. Details of this network can be found in the supplementary material of Jara-Arriagada et al. (2024).

To showcase the foundational idea and validate the proposed problem formulation and algorithm, all simulated pipes in the network were assigned uniform material properties in this study. As a result, the wave speed remained consistent across the entire network. Furthermore, in all scenarios, the same arrangement of hypothetical sensor nodes M was applied unless indicated differently. The quantity and placement of these hypothetical sensors within the network were chosen to strike a balance between comprehensive spatial coverage and maximising distinct transient paths. These sensor nodes, represented with black markers in every figure, are the nodes from which pressure measurements were obtained to determine t, the times of arrivals of pressure waves at the assumed measurement locations. The description of the tested scenarios is presented below.

Description of scenario 1

In this scenario, a pipe burst is simulated in node 159 (see Figure 8.2) which generates the transient pressures at the sensors nodes presented in Figure 8.3a. The burst is defined in TSNet with a start time at 1 s, a time for the burst to fully develop of 1 s, and a final burst coefficient of 0.01 $(m^3/s)/\sqrt{m}$. In this scenario, the network model used in the formulation of the transient localisation problems (8.4), (8.5) and (8.7) is the same as the network model used for the transient simulations, i.e., there is no pipe blockages in the

network.

This scenario was designed to investigate the effectiveness of the problem formulations (Problem (8.4), Problem (8.5) and Problem (8.7)). Problem (8.4) presents the case of reformulating the scoring algorithm into an optimisation problem that uses the fastest path in a network to find the source of a transient. Problem (8.5) extends this formulation to consider k-fastest paths, in order to assess cases where the network model does not agree with the actual model present in reality. Problem (8.7) was developed in order to improve the selection of wave propagation paths by the incorporation of weight function into the problem formulation. The focus of this scenario is to show the need for this weight function. For comparison, results from the scoring method from Equation (8.1) are presented for this scenario and all other scenarios.

Description of scenario 2

The second scenario aims to investigate the performance of Problem (8.7) in the presence of connectivity uncertainty (e.g. a closed valve). A pressure transient caused by a pump shut-off operation is simulated using a modified network model in which one pipe is removed to represent a pipe blockage (see Figure 8.2). The pump shut-off operation is simulated in TSNet at the PUMP1 in the network model (next to node 61 in Figure 8.2). The parameters used in the pump shut-off were: a pump closure start time of 1s, a pump closure period of 1s, an end open percentage of 0, and a closure constant of 1. The resulting pressure transients measured at the sensor nodes are presented in Figure 8.3b. The full network (with all pipes, including the removed pipe) is used to formulate the optimisation problem (8.7) in order to test its applicability to accurately identify the source and paths of the pressure wave to the sensors.

Description of scenario 3

The third scenario considers the same pipe blockage and the same purpose as scenario 2 for testing Problem (8.7). However, in this case, the transient source is a large customer demand in node 159 (See Figure 8.2). This combination of the transient source and the pipe blockage causes the transient pressure waves to travel along very high order fastest paths, in contrast to scenario 2. That is, the blockage in scenario 2 causes the transient waves to travel up to a 3rd fastest path for some sensors (see the results section). In scenario 3, however, the pipe blockage is in such a position that the pressure waves are forced to 'go around' the network in order to reach to the sensors in the upper part of the network as presented in the results.

The main reason to showcase scenarios 2 and 3 is to present that, in some cases, as in scenario 2, an unknown pipe blockage might have little impact on the performance of a transient source localisation algorithm (see benchmark comparison in the results section). However, in scenarios such as 3, the impact is large and, therefore, the need for the formulation of Problem (8.7) to assess networks with uncertain connectivity. In addition, scenario 3 also offers the possibility to clearly identify the portion of a path where a potential blocked pipe exists.

The large customer demand is simulated with the following parameters in TSNet: a total demand period of 20s, a demand pulse start time of 1s, a demand pulse transient time of 0.2s and a demand peak amplitude of 5 (dimensionless). As with the other two scenarios, this scenario is benchmarked against the transient source scoring algorithm presented in Equation (8.1).

Additional testing scenarios

In addition to the previous three scenarios, 8 more scenarios were evaluated to show the applicability and robustness of the problem formulation to other situations. These scenarios are described in Table 8.2. Scenarios 4 to 11 are presented in the Appendix B

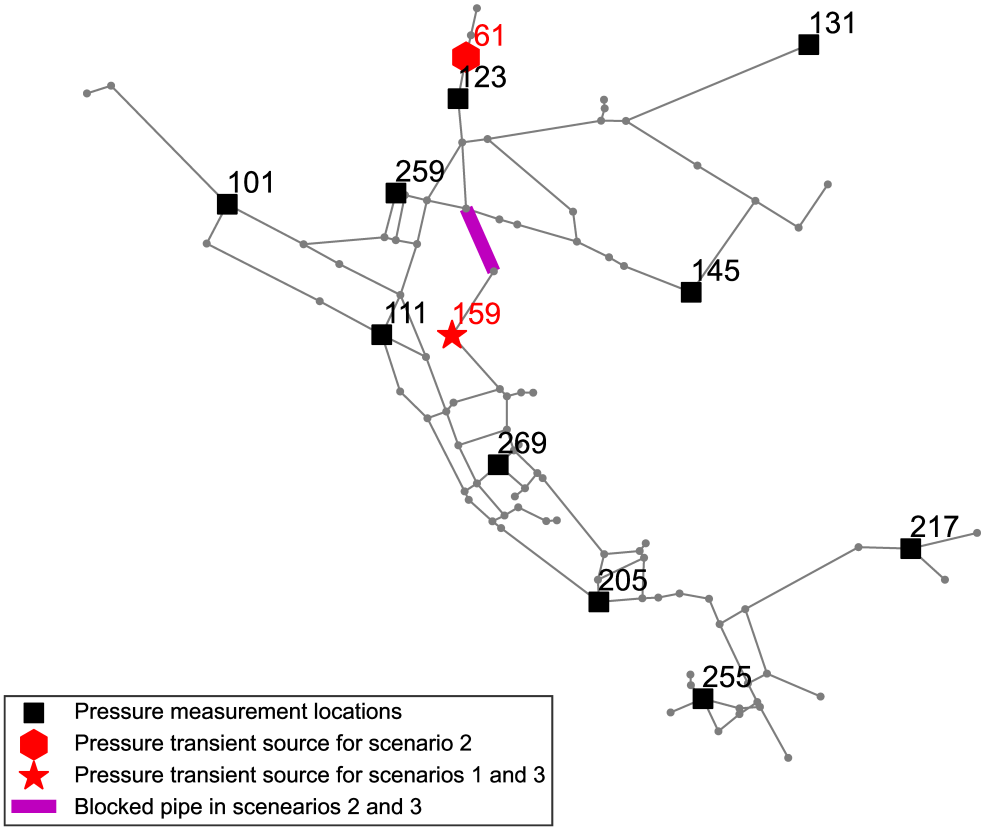
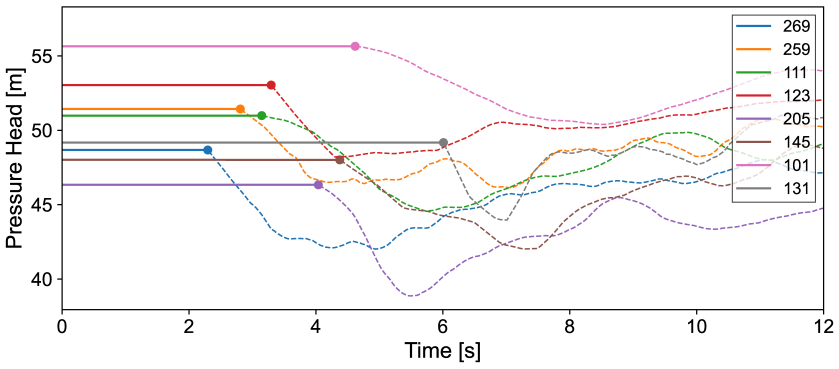
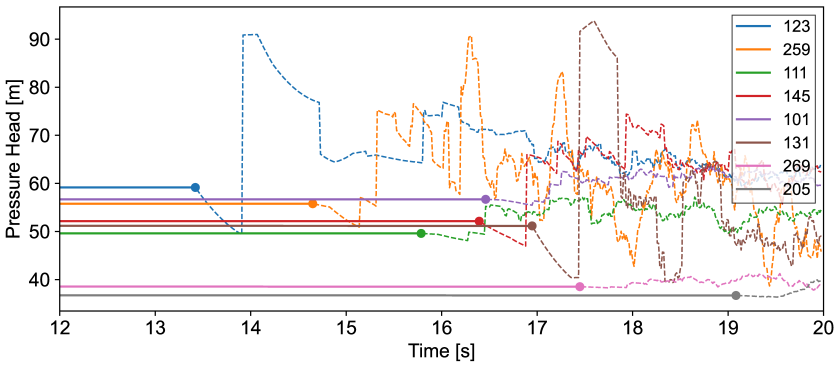


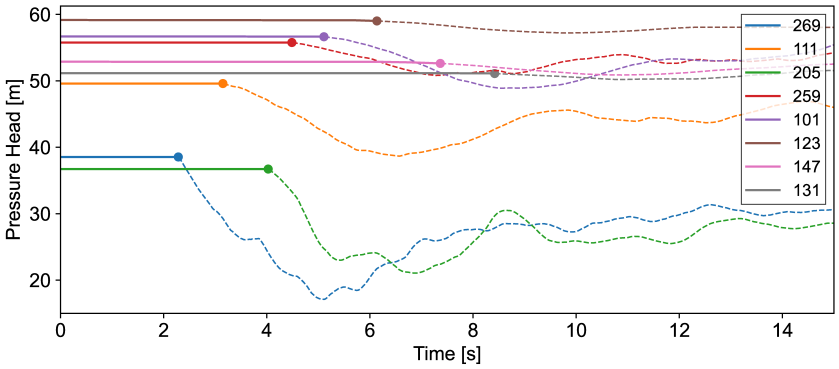
Figure 8.2: Network net3 used for testing the transient source localisation algorithm. Squared nodes are the locations of the fictitious sensors used to track pressure transient waves in the network. From: Jara-Arriagada et al. (2024).



(a)



(b)



(c)

Figure 8.3: Time series of pressure transient for the presented scenarios. a) Time series of transient waves detected by sensors in scenario 1. b) Time series of transient waves detected by sensors in scenario 2. c) Time series of transient waves detected by sensors in scenario 3. From: Jara-Arriagada et al. (2024).

section of this thesis. The performance of the proposed transient source localisation method in the additional scenarios was assessed by the number of nodes away from the true source. Utilising other metrics, such as distance from the true source, was deemed unsuitable for the first part of this study. This is because the presented transient source localisation formulation during the stage of evaluation using simulated data worked on the initial network provided without further discretisation of the network branches. This means that the set of potential solutions only rely on the available nodes in the network. Therefore, if a found solution is not in the exact source node, the distance of this solution to the source node will depend on the pipe length. This makes distance unsuitable for comparison, and for assessment of performance of the proposed method. While using the number of nodes away as a metric might also have its limitations, it is considered adequate for demonstrating the algorithm's proof of concept.

Pressure transient onset detection

Detecting the onset of pressure transients involves pinpointing the exact moment a pressure wave first reaches a pressure sensor. Several techniques are designed for this task, and their applicability often depends on the degree of noise (like minor transients echoing within the system) accompanying the wave, as evidenced by references such as (Sriramgarajan et al., 2010; Misiunas, Lambert, Simpson & Olsson, 2005; Misiunas, Vítkovský, Olsson, Simpson, Asce & Lambert, 2005).

In the first part of this study, pressure transient data were obtained from a simulation software, with no noise introduced to the pressure time series data. Consequently, the start of a transient wave at a sensor was marked when the gradient of the pressure signal in the time series crossed a predetermined threshold of -0.0045, a value derived through empirical observations. For the field study validation, additional criteria were required in order to the detect the onset of pressure transients. These criteria are presented in Section 8.4.2.

8.4 Field case study validation

To further validate the presented transient source localisation method, the method was tested using high resolution water pressure data acquired from a real operational water distribution network. This section summarises the preparation of the case study, the procedure used to assess the performance of the transient source localisation method, and the influence of the number of sensors for this particular case study.

8.4.1 Case study overview

In this case study, a large industrial user (a soft drinks manufacturer) repeatedly generated pressure transient events by operating a pneumatically actuated gate valve on a 200-mm ID pipeline to fill an elevated water tank. Six time-synchronised, high-resolution pressure monitoring devices were deployed across the network to continuously record water pressure (Figure 8.4). The monitoring campaign spanned two weeks, capturing data at 64 samples per second [S/s].

Pipe materials and installation dates, together with pipe diameter data compiled from the network's Geographic Information System (GIS) and hydraulic model, were used to estimate pipe wall thicknesses based on the relevant British Standards for the corresponding installation period. For simplicity, potential corrosion effects on reducing wall thickness were not considered. Table 8.1 presents the specific values for the modulus of elasticity and Poisson's ratio, as well as the references for determining wall thickness. These physical properties were then used to compute pressure transient wave speeds (a) in each pipe using Equation (1) (Ellis, 2008).

$$a = \sqrt{\frac{K/\rho}{1 + D/s \cdot K/E \cdot c1}} \tag{8.9}$$

This equation is derived from the fundamental principles of mass and momentum conservation for flow in a cylindrical conduit. It depends on the bulk modulus of elasticity

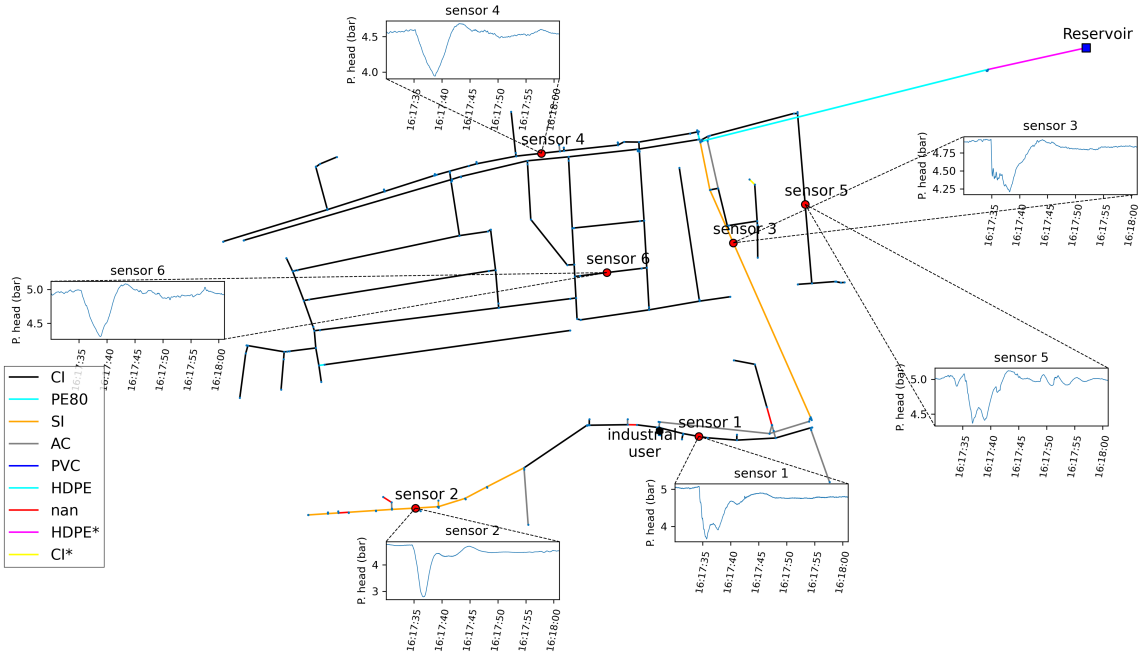


Figure 8.4: Network model and sensors' locations. Pipe materials are depicted with different colours. An example of the transient wave detected by every sensor is presented in the boxes coming out from the sensor locations.

of water (K), the density of water (ρ), the pipe diameter (D), pipe wall thickness (s), modulus of elasticity of the pipe (E), and a constraint factor ($c1$). This constraint factor , $c1$, varies according to the restraint conditions of the pipe and its thickness-to-diameter ratio. For this analysis all pipes are assumed to be anchored at both ends, and the following formulae are used for $c1$ (Ellis, 2008); a) case of thin walled pipes ($D/s \geq 25$): $c1 = 1 - \mu^2$. b) case of thick walled pipes ($D/s < 25$): $c1 = \frac{D}{(D+s)}(1-\mu^2) + \frac{2s}{D}(1+\mu)$, where μ represents the Poisson's ratio of the pipes.

Table 8.1: Pipe material properties used to determine pipe wave speeds in the network.

Material	Modulus of Elasticity (GPa)	Poisson ratio	Source for pipe thickness
Cast Iron	90	0.25	BS78-1017
Spun Iron	150	0.25	BS1211-1958
PVC	3.3	0.25	BS3505-1986
PE80/PE100	0.7/1.0	0.46	BS12201-2-2011
Asbestos Cement	24	0.3	BS486-1981

*BS stands for British Standard

8.4.2 Pressure transient onset detection

Pressure transient events were continuously monitored for two weeks, with particular emphasis on determining the precise onset of wave arrival times. This onset refers to the first moment a pressure sensor detects a transient event. The observed pressure fluctuations (Figure 8.5), marked by both rising and falling pressures, were attributed to the operation of a pneumatically actuated control valve by an industrial user (confirmed through a site visit).

Various algorithms for detecting transient onsets have been proposed (Hampson, 2014), employing wavelet analysis, Hilbert transforms, gradient detection, and probabilistic methods (Hoskins, 2015; Hampson, 2014). Hampson (2014) demonstrated that a gradient-based detection strategy, relying on the maximum gradient change in the pressure signal, offers a favourable balance of accuracy and consistency. After preliminary trials, a slightly modified version of this gradient-based approach was developed , given its simplicity and minimal computational requirements. This algorithm could also be embedded within the firmware of battery-operated pressure monitoring devices, thereby reducing the need to transmit high-resolution pressure data wirelessly.

Hundreds of transient events were available for analysis. Although small, a notable distinction emerged between rising and dropping transients at the front wave arrival; rising transients exhibited a more gradual increase in pressure, whereas dropping transients were typically sharper. This disparity likely results from the speed at which the industrial user opens or closes the valve. Because sharp signals facilitate more accurate detection of the wave arrival onset, the evaluation of the proposed source localisation method focused on the dropping transients (associated with valve opening). Moreover, a set of heuristics was developed to segregate transient events from the long pressure time series (Figure 8.5a). A detailed illustration of the onset detection process provided in Figure 8.5b.

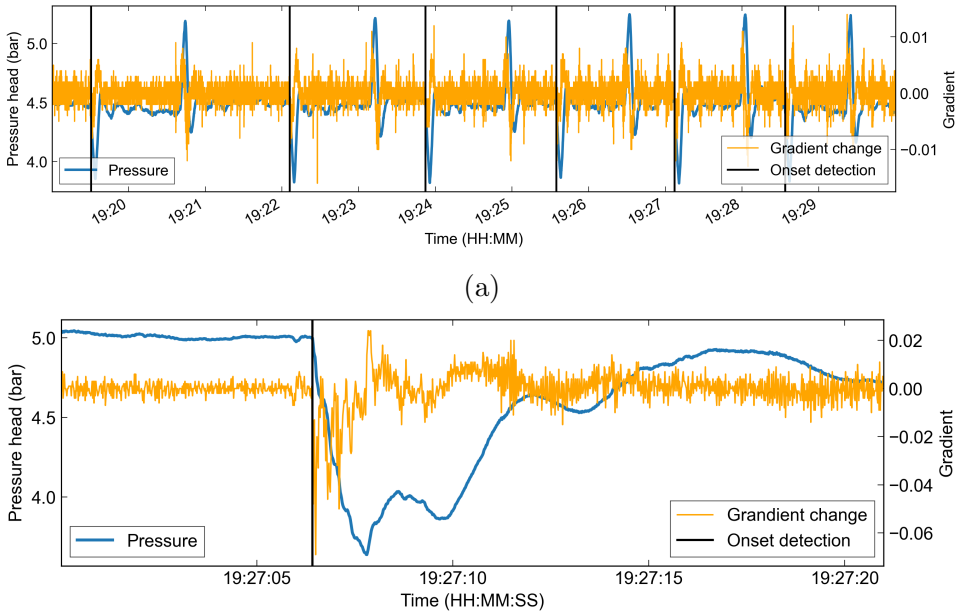


Figure 8.5: Detection of dropping transient pressure onset arrival time using the gradient method. a) multiple transient detection with the aid of heuristics. b) Detail view of transient pressure onset detection.

8.4.3 Evaluation trials and estimation of errors in the solution

As an initial test, the method was applied on a well-defined transient event recorded at approximately 03:00 AM, when consumer demand and routine operational activities were minimal, thereby reducing the dampening effects on the pressure transient and improving the accuracy of transient onset detection. To quantify the method's error, a radius of approximately 150 m was discretised around the transient source with additional nodes spaced at 20 m intervals. This spacing was selected through trial and error to balance the need for a more refined solution against the increased computational requirements of adding more nodes. In this case study, the source of the transient events was known to be an industrial user located at a specific location. This site was assigned to a terminal node of a valve in the network and designated it as the true source location. The distance between the true source node in the network and the node identified by the transient source localisation method was determined via the shortest path. This distance was taken as the error for the transient source localisation method.

As detailed in Section 8.2.1, the method relies on the selection of a factor α, that slightly

increases the calculated time of wave propagation in the path selected by the algorithm. A value of 0.06 was selected for the analysis of the real case study based on the obtained results for the simulated pressure transient scenarios.

In addition to the initial trial test, multiple analyses on pressure data were performed that included 93 transient events in total. The primary objective was to characterise the distribution of potential errors in transient source localisation, given the variability in wave arrival times identified through the transient onset detection method.

8.4.4 Influence of the number of sensors

Analysing the influence of the number of sensors is a time consuming task as there are 15 possible combinations for removing one sensor at a time from six sensors until only 2 sensors remain. For simplicity, this section only analysed the influence of removing the farthest sensors away from the source until only two sensors are left for analysis. For every set of sensors transient source localisation were performed for different times of the day. The results of these multiple experiments are then visualised in box plots to assess the performance of the method for transient source localisation against the number of sensors in this particular case study.

8.5 Results

This section presents and discusses the results of the scenarios proposed to evaluate the presented transient source localisation method.

8.5.1 Scenario 1

In this scenario, a basic transient source localisation case was used to evaluate the performance of the problem formulations described in Problem (8.4), Problem (8.5) and

Problem (8.7). Figure 8.6a shows the benchmark solution for transient source localisation obtained with the scoring algorithm (Equation (8.1)) presented in Srirangarajan et al. (2010) and Zecchin et al. (2022). The scoring algorithm correctly identifies the transient source. The score difference among the potential source nodes close to the solution is relatively small, however. In comparison, the solution obtained by formulating the transient source localisation as a regularised continuous optimisation problem (Problem (8.4)) is illustrated in Figure 8.6b. The correct solution in this case is clearly separated in score from other potential nodes (sources), and the regularisation term enables selection of the size of the search area for the most likely source. Figure 8.6c shows the results of transient source localisation by extending the optimisation approach to include k-fastest paths for the pressure wave propagation from the source to all sensors (Problem (8.5)). The optimisation approach selects the k- fastest path that minimises the objective function. Figure 8.6d shows the results of the transient source localisation with the addition of a weight function (Problem (8.7)). The weight function (Equation (8.6)) aims to address the potential issue of the algorithm selecting an incorrect propagation path that matches the times of arrival due to differences in the analytical results with measured values.

The legend of Figure 8.6c and Figure 8.6d displays the order of the chosen path between sensors and the source in squared brackets, and also the corresponding error value between the simulated and measured times of arrival, which was defined in Equation (8.8). Additionally, the colour scheme used indicates the individual pressure wave paths to each sensor from the source in the network. The solution obtained in Figure 8.6d, shows that the algorithm correctly identified the source of the transient and the fastest paths between the source and every sensor. However, in Figure 8.6c, where the weight function was not applied, the algorithm attempted to compensate for the error from the uncertainties in wave arrival times by incorrectly selecting a 2^{nd} and 3^{rd} fastest paths to the sensors 131 and 101, respectively, which were not necessary as there was no blockage in this case. This demonstrates the importance of using the weight function.

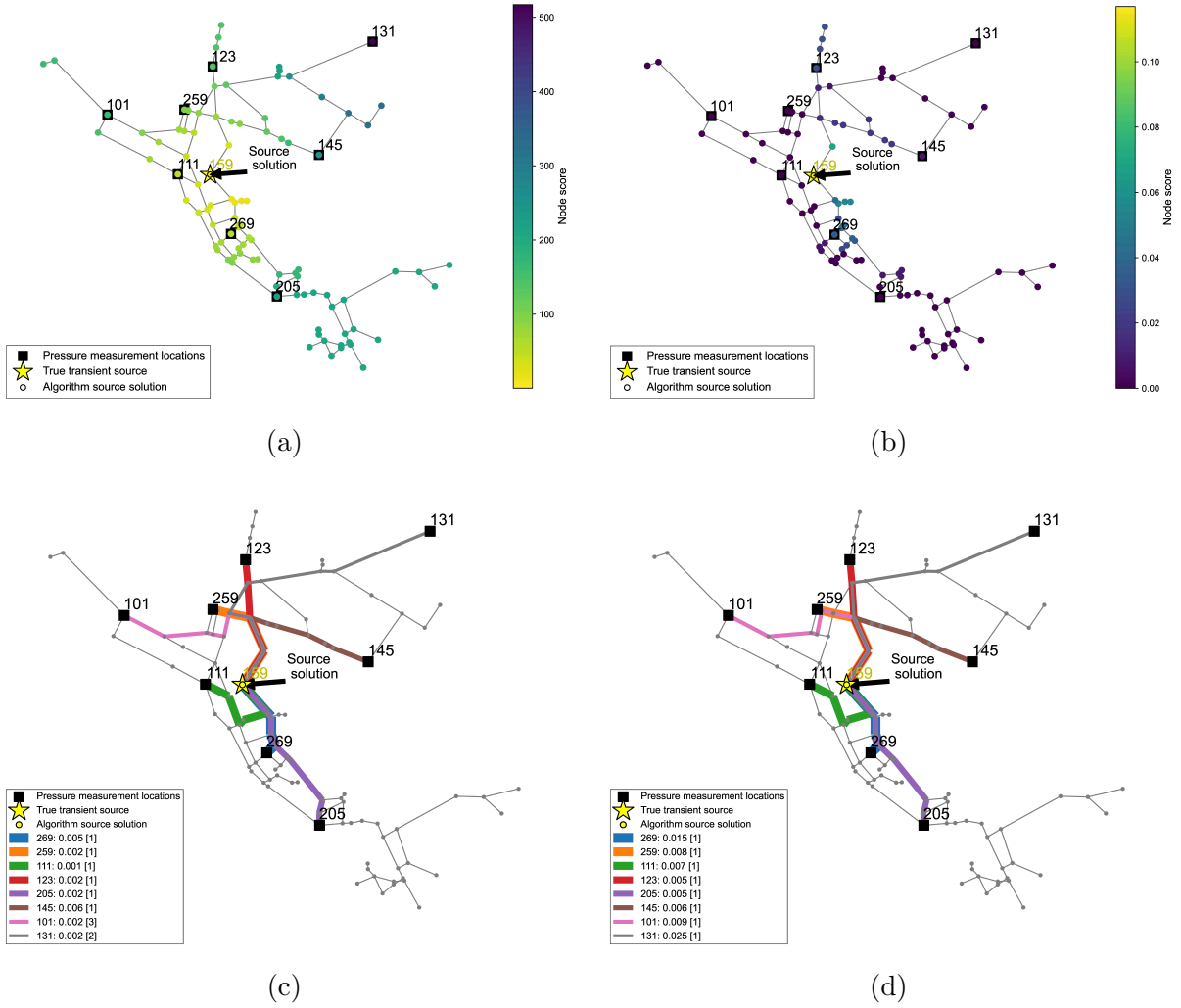
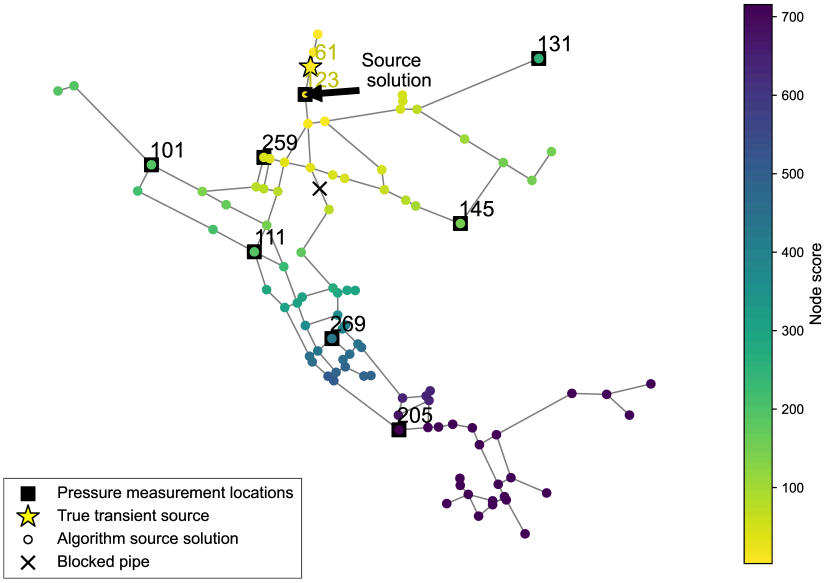


Figure 8.6: Solutions for the scenario 1 for transient source localisation by different algorithms. a) Benchmark scoring algorithm solution for transient source localisation in scenario 1. b) Transient source solution for scenario 1 by conversion of the scoring algorithm into an optimisation problem with regularisation. c) Transient source solution for scenario 1 by applying the developed k-fastest paths transient source localisation algorithm without weight function. d) Transient source solution for scenario 1 by applying the developed weighted k-fastest paths transient source localisation algorithm. From: Jara-Arriagada et al. (2024).

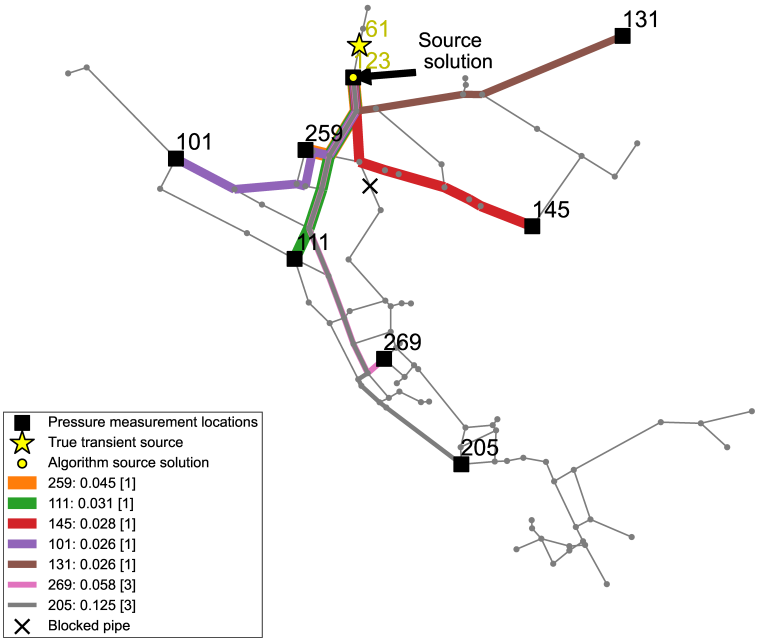
8.5.2 Scenario 2

In this scenario, a transient simulation was performed after removing a pipe from the network model to simulate a valve that was unknowingly closed. The removed pipe is represented by a thick line in Figure 8.2 and indicated by a cross in Figure 8.7. Because of the removed pipe, transient waves had to propagate through alternative paths. These alternative paths correspond to the 3rd fastest paths to sensors 269 and 205 in the lower portion of the network (assuming that the full network without pipe removal was used by the transient source localisation algorithm). Under this scenario, the propagation time differences between the 1st and 3rd fastest paths were minor. As a result, the scoring algorithm (Equation (8.1)) used for benchmark was able to accurately identify the potential location of the transient source (Figure 8.7a). However, the benchmark scoring algorithm used in Figure 8.7a is not able to provide information about the blockage in the network, and only identifies the potential location of the transient source.

The solution of the proposed k-fastest path problem formulation (Problem (8.7)) is illustrated in Figure 8.7b. The solution obtained correctly identifies the transient source and the paths to the sensors taken by the transient wave, which provides valuable information about the hydraulic connectivity of the network. In particular, Figure 8.7 clearly shows that the pressure wave avoided the path that included the removed pipe in the network to sensors 269 and 205. This indicates the presence of a closed valve or blocked pipe along this path. Given this information, then, a water operator can look into the network records for any nearby valve that might be closed in the identified path. Additionally, if no records exists, the information provided by the algorithm can be used to perform a field exploratory campaign to check for any potential unreported closed valve or pipe blockage.



(a)



(b)

Figure 8.7: Solution for the scenario 2 for transient source localisation. Optimisation approach for multipath transient source localisation vs scoring algorithm. a) Benchmark transient source solution (scoring algorithm) for scenario 2. b) Weighted k-fastest paths transient source localisation algorithm solution for scenario 2. From: Jara-Arriagada et al. (2024).

8.5.3 Scenario 3

In this scenario a pipe was removed to simulate a closed valve, as in scenario 2, which deviates the flow through the pipe. The transient source was set at node 159, see Figure 8.2. In this scenario, transient waves had to travel through significantly longer and alternative paths to reach the sensors in the top half of the network. This scenario shows the benefits and limitations of the proposed transient source localisation method (Problem (8.7)).

The proposed transient source localisation method had to explore high rank fastest paths to find a stable solution for the problem. This is because many low rank k-fastest paths are minor variations of the fastest paths in this scenario. The solution for the transient source obtained in this scenario is one node away from the true solution (see Figure 8.8b). However, this solution shares the same paths as the true solution for all sensors. As explained by Zecchin et al. (2022), this occurs because the algorithm analyses differences in times of arrivals of the pressure waves to pair of time-synchronised sensors. If two sources result in overlapping paths, then the difference in times of arrivals is the same and therefore the two sources have the same likelihood to be the true source. In the case of the provided solution, the paths between the true source (node 159) and all network sensors overlap with the paths between the identified source (node 161) and the sensors. Nevertheless, the provided solution is sufficient to inform a water operator of the potential source of a transient if an exploratory campaign is performed around the identified source.

In comparison to the solution from the optimisation approach, the benchmark solution is 6 nodes away from the true source, and many potential source nodes share similar scores (see Figure 8.8a). This highlights the advantages of the proposed problem formulation. Additionally, as in scenario 2, the scoring approach does not provide any connectivity information. In contrast, the solution of the optimisation problem (8.7) clearly shows that a path in the network was avoided by the pressure waves (see Figure 8.8b), signifying that a blocked pipe exists in the network. In fact, for any transient source scenario, if the solution of the optimisation approach explores more than a fastest path to any sensor, this is an indication of a potential blockage in the network. This assumption is reinforced

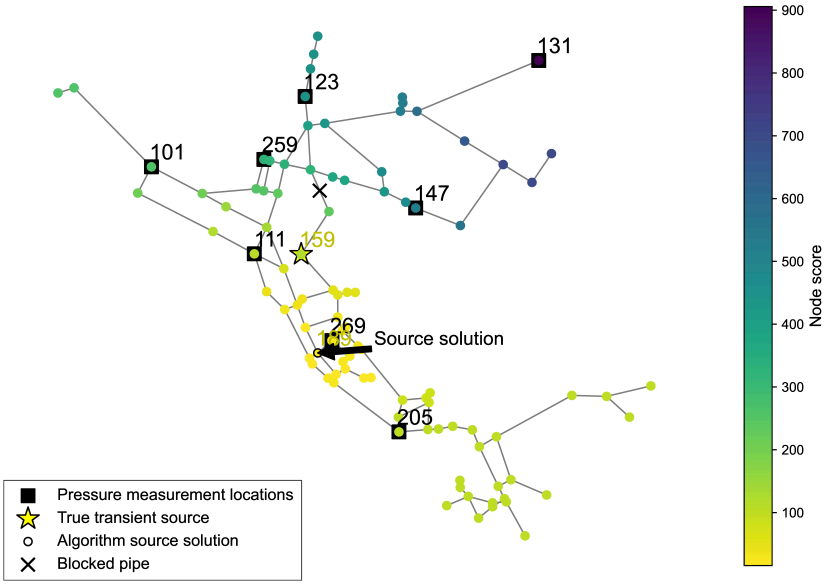
if the blocked pipe is avoided in the paths between the source and multiple sensors.

The legend in Figure 8.8 displays in brackets the rank of the paths taken by the pressure wave between the source and the sensors. The selected paths to sensors 269, 111, and 205, are the (1st) fastest paths, which is accurate as the simulated pipe blockage is not located on the fastest paths between the source and those sensors. It is observed that higher ranked paths are selected between the source and the rest of the sensors. These paths are selected because the paths that included the blocked pipe generate times of pressure wave arrivals shorter than the actual times that the pressure wave took (the transient simulation was performed excluding a pipe). However, although these paths accurately avoided the blocked pipe, the solution of the optimisation approach still compensated some errors to minimise the objective function by selecting slightly higher rank fastest paths than required. A clear example of this is the path to the node 147. From the start, the path between node 161 and node 147 (in pink) deviates from all other paths between 161 and the upper part of the network, adding extra travel time. The path to node 147 then merges with the other paths, but it deviates slightly later on, passing through node 111 when it was not required to. The fact that the path to node 147 passed through pipes not considered in the the share route of this path to other sensors indicates that the path to node 147 might be overcompensating errors.

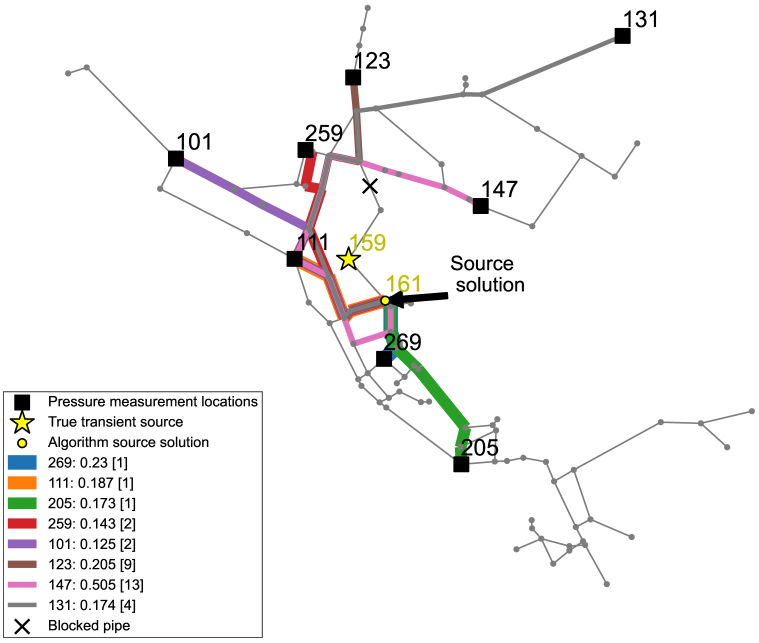
Overall, the solution provided by the optimisation approach is satisfactory, indicating a solution very close to the true source and the path containing a potential blockage. The challenges and limitations of the presented approach rely on accurately determining the pressure wave arrival times and the value selected for the constant α in the weight function.

8.5.4 Additional testing scenarios

This section presents the results of additional testing scenarios performed to validate the proposed problem formulation. In general, from the tested scenarios, when no pipe blockage exists, the algorithm accurately finds an appropriate solution without considering



(a)



(b)

Figure 8.8: Solution for scenario 3 for transient source localisation. Optimisation approach for multipath transient source localisation vs scoring algorithm. a) Benchmark (score algorithm) transient source solution for scenario 3. b) Transient source solution for scenario 3 by applying weighted k-fastest paths transient source localisation algorithm. From: Jara-Arriagada et al. (2024).

higher rank fastest paths. In other words, higher ranked fastest paths were only considered when a pipe blockage existed, which provides confidence in the algorithm for detecting potential inconsistencies between an hydraulic model and the actual WDN.

In cases where blocked pipes exist, the algorithm selected paths from the source to the sensors that in many cases avoid the blocked pipe. However, there are instances in which, the algorithm avoids a certain path for a large majority of sensors but not for all of them. An example of this occurred in scenario 10. In this scenario two pipes close to the transient source are removed (Figure B.7). For this particular case, the solution (Figure B.7a) was capable of considering the two blockages to the paths of all sensors, except to sensor 101. The algorithm was applied several times, adjusting the value of α, to test if another value would help the algorithm adjust its solution. However, the solution remained stable, with the algorithm selecting the blocked pipe for the path to node 101.

In order to attempt to reach to a better solution for this scenario, sensor 101 was removed. The results of removing sensor 101 are presented in Figure B.7b. In this case, the solution correctly identify the two blocked paths, the transient source and the paths to the sensors. This result is interesting and encourages further research in order to understand the effects of the placement and number of sensors.

Table 8.2: Validation scenarios for the solution of Problem (8.7). From: Jara-Arriagada et al. (2024).

Scenario	True transient source node	Blocked pipe	α	Source solution	Nodes apart	Blocked path found	Highest rank fastest path
4	PUMP2	-	0.06	Lake	0	-	1
5	PUMP2	117	0.06	101	1	Yes*	7
6	119	-	0.07	119	0	-	1
7	119	123	0.07	119	0	Yes	2
8	149	-	0.05	149	0	-	1
9	149	177	0.05	173	1	Yes	5
10	PUMP1 (61)	122 & 129	0.06	123	1	Yes**	3
11***	PUMP1 (61)	122 & 129	0.06	123	1	Yes	2

PUMP 2 is located between lake and node 10.
* The blocked path is avoided by all sensor's paths except for the path to sensor 131.
** The blocked path is avoided by all sensor's paths except for the path to sensor 101.
** Considers same conditions of scenario 10 but with the removal of sensor 101.

8.5.5 Field case study validation results

This section summarises the results obtained after analysing the high-resolution pressure data and implementing the method for transient source localisation on the operational network used as a case study. The results are summarised in three subsections as follows.

Performance transient source localisation method

Using the gradient detection algorithm described in Section 8.4.2, the wave arrival onset times were derived for each sensor from the collected high-resolution water pressure time series. To conduct an initial evaluation of the transient source localisation method, one transient event was selected from the dataset. Particular care was taken to accurately determine the wave arrival onset times, thereby minimising potential errors. The corresponding localisation results are shown in Figure 8.9.

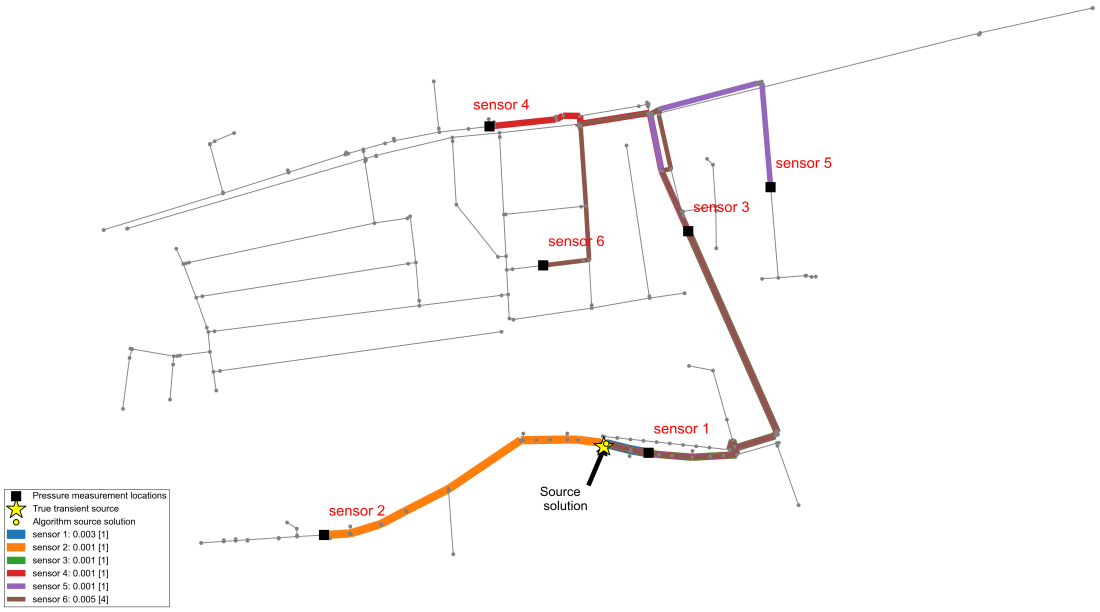


Figure 8.9: Results of the pressure transient source localisation method, on a transient event recorded near 03:00 AM

The transient source localisation results for this scenario display a high degree of accuracy, with an estimated error of approximately 8 m. This accurate result is attributed to

the relatively low noise level in the pressure signals recorded during the transient event. Notably, the algorithm selected the fourth-shortest path for Sensor 6, suggesting a possible network blockage or, alternatively, an inaccurate estimate of the pipe wave speed along that particular route.

Accuracy on wave arrival time onset detection and impact on localisation error

Multiple transient events were examined for this analysis. Figure 8.10 illustrates the distribution of wave arrival times at different sensors in the network, relative to Sensor 1. The shape of these distributions primarily reflects the errors arising from the gradient detection approach rather than fluctuations in pipe wave speeds, although the latter may also contribute to the observed variability. This variability highlights the uncertainties that any method for transient source localisation must address and serves as a benchmark for evaluating other transient onset detection techniques.

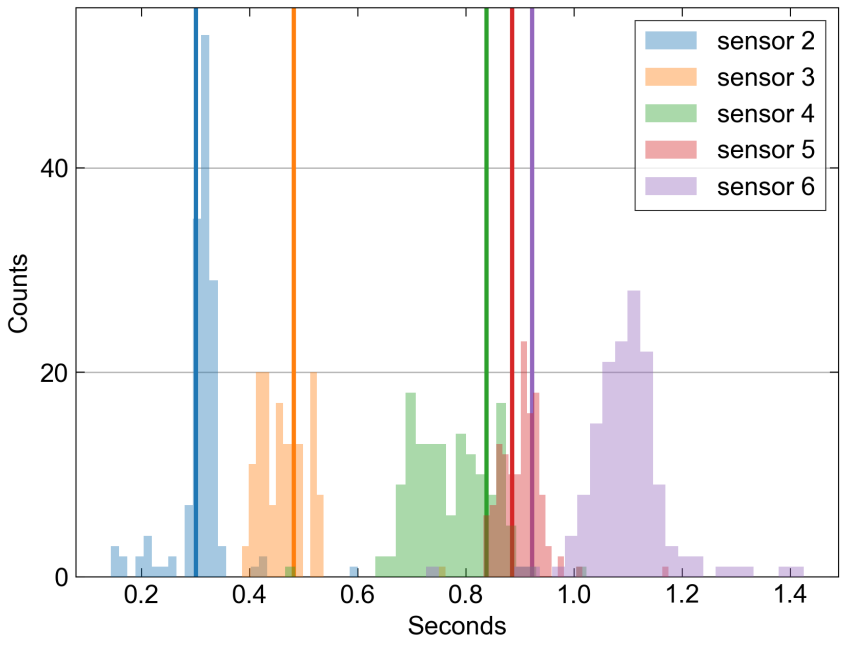


Figure 8.10: Distribution of pressure wave travel times with reference from Sensor 1. Vertical thick lines represent the estimated value of travel times based on predetermined pipe wave speeds and pipe lengths.

Based on Figure 8.10, the transient onset detection algorithm performed particularly well for Sensor 2, whereas sensors 3 and 4 did not exhibit a well-defined distribution of arrival times. Figure 8.10 also includes thick vertical lines representing the analytically determined arrival times derived from the shortest-path distance between Sensor 1 and each of the remaining sensors. These calculated travel times correspond reasonably well to the observed wave arrival time distributions, except for the path to Sensor 6. As outlined in Section 8.5.5, the transient source localisation algorithm identified the path to Sensor 6 as potentially having a blockage or being subject to larger errors in the estimated pipe wave speed.

The same transient events used in generating Figure 8.10 were analysed to locate the source of the pressure transients. The error distribution obtained by the transient source localisation method using the full set of six sensors is displayed in Figure 8.11. This distribution is approximately normal, with a mean of 44.7 m, a median of 48.53 m, and a standard deviation of 19.98 m. These values closely match previously reported localisation errors, where the median error was 43.79 m (Srirangarajan et al., 2013). The results from multiple transient events demonstrate that the method can be applied in operational networks despite potential uncertainties related to wave arrival time detection, pipe wave speeds, sensor locations accuracy, and pipe lengths.

Impact of the number of pressure sensors on localisation results

In this analysis, the sensors located farthest from the source were consecutively removed , one by one. A total of 23 transient events recorded over a single day were examined to characterise how reducing the number of sensors influences the resulting error distributions. While a transient simulation could have been used to assess the impact of the number of pressure sensors, it was chosen to analyse the acquired pressure data in order to capture the influence of uncertainties in wave arrival times and assess how fewer sensors contend with those uncertainties. Figure 8.12 shows how the error distribution evolves as each farthest sensor is removed.

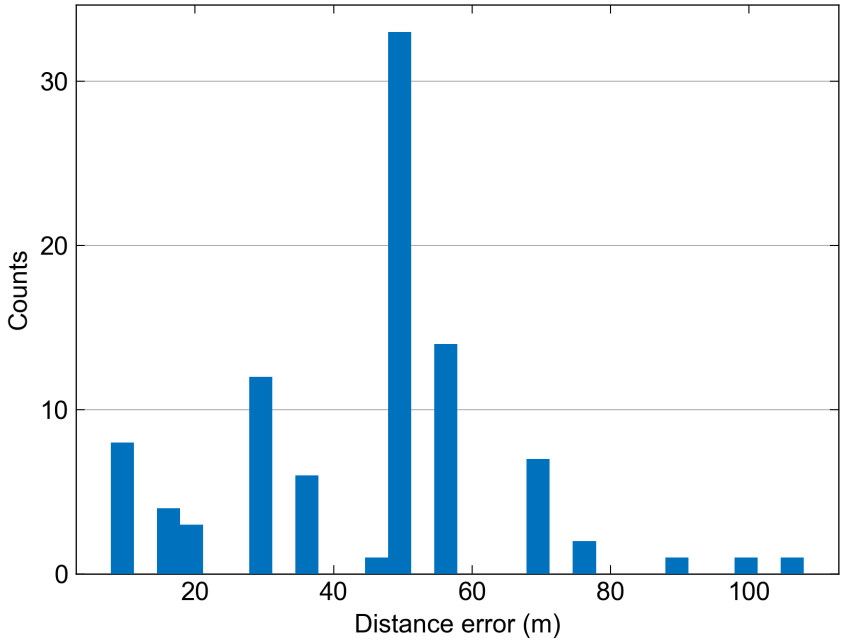


Figure 8.11: Distribution of source localisation errors.

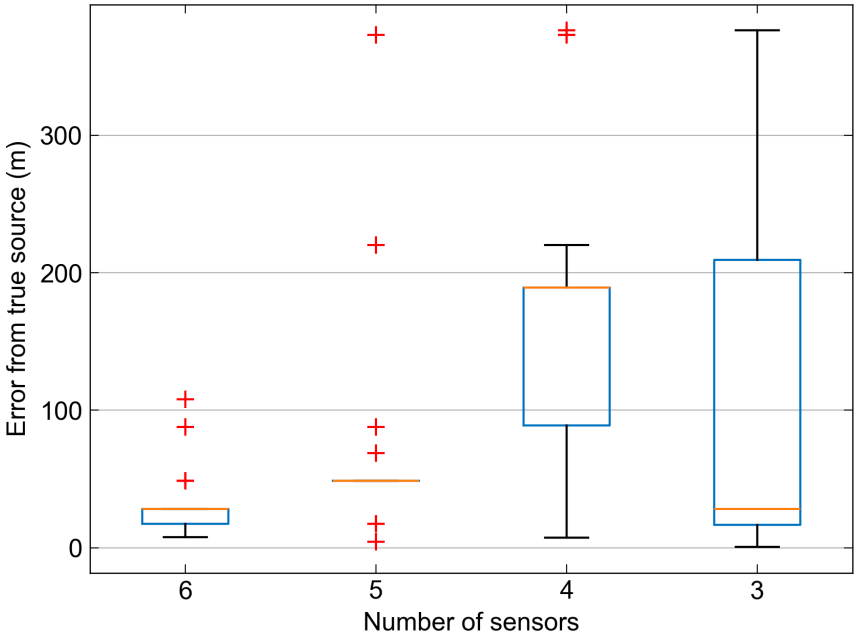


Figure 8.12: Influence of number of sensors in source localisation errors.

8.6 Conclusions

This chapter presented and investigated a novel approach for pinpointing transient source locations, leveraging graph theory, k-fastest paths, and disparities in the arrival times of pressure transient waves at sensors within water distribution networks. This represents a pioneering method addressing transient source localisation in networks with uncertain connectivity. Exploration of the literature shows that the problem of transient source localisation in networks with uncertainty in connectivity has never been previously studied.

Through the formulation of an optimisation problem that considers the most likely location for the source of a transient wave and the k-fastest paths that minimise the differences in arrival times between sensors and analytical times of arrivals, a robust algorithm has been developed that can accurately identify the source of a transient event, as well as the potential path that the wave may have taken, even in cases where the hydraulic model does not accurately capture pipe hydraulic connectivity. The algorithm's performance was evaluated through various simulated scenarios, and a real field case study to test its efficacy in transient source localisation.

The algorithm demonstrated good performance in all simulated tested scenarios, particularly in scenarios where blockages lead to an important increase in pressure wave propagation times. In such scenarios, the proposed formulation in this chapter outperformed a previous scoring-based transient source localisation algorithm from the literature. The scoring-based algorithm failed to identify the correct source of a transient event in the presence of such a blockage, while the proposed approach succeeded and was able to identify the path taken by the pressure wave. This allows for the detection of potential unreported pipe blockages or closed valves in the network when the selected path deviates from the expected fastest path. This advantage provides a new use for differences in times of arrival of pressure waves, complementing other algorithms for localising pipe blockages, such as the one presented by Pecci et al. (2020).

The robustness of the algorithm for finding pipe blockages is also confirmed by its per-

formance when no pipe blockage exists. Practically, in all simulated scenarios that were tested without a pipe blockage, the algorithm found an appropriate solution without relying on higher rank fastest paths. This adds confidence that when the solution selects a higher rank fastest path, it is likely that there exists a blocked pipe in the network. This claim, however, depends on an appropriate calibration of the value of α in the weight function. For all scenarios, a value of α of 0.06 provided the best general performance.

In cases when there is a pipe blockage in the network, most of the paths between the source and the sensors selected by the proposed solution method accurately reflect the presence of a blockage. However, there are instances in which, for a particular sensor, the blocked path is also considered as part of the solution. Despite this drawback (see, e.g., scenario 10), there are usually a minimal number of paths that consider the blocked pipe against several others that avoid the path. This suggests that a blockage exists in the network and consequently, the provided solution in such cases is still valuable.

In the analysis of the field case study, the algorithm generally selected the first shortest path for all sensors, except for Sensor 6. A detailed examination of Figure 8.10 indicated that the predetermined wave travel time for Sensor 6 was significantly different from the measured travel time. Several sources of uncertainty can influence transient source localisation. Nevertheless, depending on the travel time of any alternative path, certain uncertainties that could have led the algorithm to select this path may be excluded.

For instance, the detection of transient onset depends partly on the sampling resolution, which in the field case study was 64 S/s. This resolution implies a maximum error of approximately 0.0161 s in wave arrival time, corresponding to around 20 m of distance at a wave speed of 1240 m/s (characteristic of cast iron). Additionally, gradual increases or decreases in pressure can introduce further error in detecting the arrival time, which can reach up to 0.1 s (or 124 m) for Sensor 6, as illustrated in Figure 8.10. However, based on the transient source localisation method and the observations in Figure 8.10, the alternative path selection suggests that the algorithm is compensating for a larger time increment in the wave's travel path. Sensor location errors could also affect the

solution and either compound these errors or serve as an independent source of error. Nevertheless, in the particular analysed field case study, the pipeline segment hosting the sensor is just over 100 m in length (equivalent to about 0.08 s of wave travel in cast iron), making it less likely that an alternative path was chosen solely due to an incorrect sensor position.

From Figure 8.10, it is observed a 0.175 s discrepancy between the analytically predetermined wave travel time for the shortest path to Sensor 6 and the measured travel time. Although it is possible to account for this deviation using the maximum expected errors discussed earlier, the measured travel time is near the mean value, indicating that it is less likely that measurement error alone accounts for the observed deviation.

These findings highlight two principal sources of uncertainty that may explain why the algorithm selected an alternative path. First, the wave speed assigned to pipes near Sensor 6 could be inaccurate. The isolated 270 m path to sensor 6 is designated as cast iron, with an estimated wave speed of around 1314 m/s, corresponding to a travel time of 0.205 s. If, however, these pipes were actually plastic, with an estimated wave speed of around 250 m/s, then the travel time would be approximately 1.080 s, introducing a large discrepancy of 0.875 s. This substantial difference could prompt the algorithm to select an alternate path if some pipes in this section are plastic. Second, the presence of a closed valve along the main path would likewise force the wave to travel along an alternative route. Although valve records for the network were not available, the algorithm's output suggests that erroneous pipe materials or valve statuses might exist in the current hydraulic model.

It is hypothesised that the number of sensors can play an important role, and a trade off in obtaining correct solutions. Few sensors can produce that the combinations of differences of times of arrivals of the pressure wave between sensors might not lead to a unique solution, and therefore, making it harder to find the correct solution. In contrast, too many sensors may potentially cause a large overall error in the algorithm as a result of the accumulated individual errors of every sensor, again, complicating the localisation of the transient source. A preliminary assessment of the effect of sensor quantity using the

field case study revealed that five sensors were sufficient to maintain robust performance. A lower number of sensors, up to four or three sensors can still be used, however, the distribution of localisation errors widens substantially.

Currently, the method pinpoints a single location as the transient source. A potential extension of this approach involves combining the k-shortest path information generated by the algorithm with an appropriate scoring framework (e.g., Srirangarajan et al. (2013)) to identify a search area for source localisation. Accurate and up-to-date information on pipe materials remains essential for ensuring robust performance of this method. Future efforts should focus on integrating the proposed approach into a comprehensive network monitoring and asset management framework, while also advancing both its computational efficiency and theoretical foundations.

Further research should also explore improvements to the weight function used in the optimisation formulation, which could better account for factors such as dissipation of transient waves and uncertainties in pressure wave arrival times. Finally, in order to test the algorithm in other operational networks with diverse pipe materials and varying sources of uncertainty, it would be necessary to determine the optimal placement of sensors to maximise the discoverability and detection of transient waves.

In summary, this chapter has presented a novel approach for localising the sources of pressure transients in water networks with uncertainty in connectivity. This method is a necessary step also for the interpolation of CPIS as presented in the previous chapter. The evaluations of the presented method have yielded encouraging results and insights. Subsequent research should focus on further evaluating the impact of sensor locations and enhancing the performance of the method in operational networks. Specifically, when implementing this method, it is essential to explore techniques to reduce the uncertainty in both estimated wave speeds and recorded arrival times. Furthermore, robust optimisation strategies should be investigated to enhance the algorithm's efficacy in the presence of these uncertainties.

Chapter 9

Conclusion

The main aim of this thesis as presented in Section 1.2 was to enhance the cumulative pressure-induced stress (CPIS) concept into three key areas of motivation, theoretical foundations and methods to support its practical application. The motivation for these areas of investigation is to provide water utilities with the main foundational requirements for the analysis of CPIS in large-scale experimental programmes. These experiments once conducted utilising the information provided in this thesis will determine if CPIS can be established as a critical metric for managing water distribution networks. To this end, this thesis has investigated a series of analytical and numerical methods so that CPIS can be properly supported and analysed in large-scale experiments. In the following sections, this chapter presents a summary of the main conclusions and contributions of the research presented in this thesis, followed by a revision of the limitations of this work and recommendations for future work.

9.1 Summary of thesis achievements

This section provides a summary of the main conclusions and achievements obtained from the research performed in this thesis. The conclusions and achievements are summarised below according to the order of the objectives presented in Chapter 1.

9.1.1 Quantification of the impact of steady-state mean pressure and pressure range with pipe breaks

From the literature review provided in Chapter 2, the first task for this thesis was to analyse the relationships of average metrics of water pressure and pipe breaks. The motivation of this task was to confirm previous findings from the literature, and to provide models to assess the influence of water pressure on pipe breaks with sensitivity analyses. These analyses were presented in Chapter 3. The models developed in this chapter using an extensive dataset of pipe breaks from a water utility in the UK effectively confirmed that water pressure is statistically related to pipe breaks. The models based on logistic regression showed that small reductions of average metrics of water pressure can lead to significant reductions on pipe breaks. In particular, models developed for cast iron pipes with breaks occurred in winter showed the most sensible impact of managing water pressure metrics for pipe break reductions.

Although the obtained results are promising, the observed magnitudes of mean pressure and pressure range do not align closely with the magnitudes required to cause a pipe to break. The data obtained for the development of Chapter 3 was curated and analysed in detail, ensuring the observed statistical evidence. Furthermore, previous findings also support the observed statistical relationships. As presented in the literature review, a mechanical explanation for the impact of mean pressure has been presented on the basis of an empirical pipe failure criterion. This approach, however, does not explain the impact of pressure range and the strong observed relationships. Reduction of pressure range as low as 5 m are shown to potentially reduce up to 43% pipe breaks. However, it is not clear the mechanisms by which this could be possible.

On the basis of the evidence of pressure fluctuations and the cumulative pressure induced stress concept, it was hypothesised that the areas observed with large pressure variation in the case study analysed in Chapter 3, might also been areas of recurrent pressure fluctuations. The main hypothesis that this thesis supports is that pressure fluctuations might weaken pipes and lead them to failure through fatigue in combinations with external

loading conditions. The evidence observed in Chapter 3 supports this hypothesis which was then tested in a theoretical analysis in Chapter 4.

9.1.2 Fracture-mechanics-based analysis of pressure fluctuation magnitudes and mean pressure levels that lead to fatigue failure in deteriorated pipes

In Chapter 4, fatigue analyses using fracture mechanics were developed to assess the conditions under which pressure fluctuations are relevant for the fatigue failure of cast iron pipes. Such analyses were needed so that CPIS is supported by theoretical foundations. The results of these analyses were supported by a comprehensive review of the fatigue assessment methods used in fracture mechanics, and the materials properties for cast iron pipes. From the developed analyses, it was concluded that fatigue failure for the case of cast iron pipes is relevant for deteriorated pipes that have suffered from extended corrosion patches reducing the pipe wall thickness. Many existing cast iron pipes in operation nowadays had installations dating back from the Victorian-era. During that time, pipes had thick wall thicknesses which aided these pipes to endure the past of time. However, a review of several pipe inspection images from a water utility revealed that various cast iron pipes observe deep corrosion patches together with cracks. Pipes under such conditions are vulnerable to fatigue failure according to the presented results.

The analysis presented in Chapter 4 had the purpose to inform the trends in changes of fatigue life by changes in the magnitude of pressure fluctuations in pipes. The analyses were performed using a representative 150 mm OD diameter distribution pipe for analysis. Therefore, the results are illustrative. The first conclusion is that the magnitude of the pressure fluctuations observed in a network can severely impact of the fatigue life of pipes. In general, a reduction of the magnitude of pressure fluctuations of 10 m can more than double the fatigue life of pipes. Mean pressure also has an important role on the fatigue life of pipes, higher mean pressure values reduce the fatigue life of pipes, which

could be related to the results of Chapter 3 with respect to the impact of mean pressure. However, the changes in the magnitude of pressure fluctuations outweigh changes in mean pressure as presented in Figure 4.20. The result obtained in this figure might explain cases observed in practice where reduction of mean pressure did not work as expected reducing pipe breaks. Usually, mean pressure reduction schemes need the utilisation of pressure control devices. If these devices are not properly installed and operated, they can increase pressure fluctuations in the system leading to increase fatigue and deterioration of pipes.

In general, the conclusions from Chapter 4 support that pressure fluctuations have an impact on deteriorated cast iron pipes. The number of these deteriorated pipes in the UK's water distribution networks might be considerable. Therefore, the option of reducing pressure fluctuations to extend the fatigue life of pipes is important. As distribution mains cannot be analysed individually, a general metric to inform the potential state of fatigue damage in pipes should be recommended for modelling pipe breaks. CPIS as metric already developed to assess the level of transient activity in water distribution networks appears suitable to inform the potential state of damage of a pipe without the need of an extensive fatigue failure analysis. Monitoring and actions to decrease this metric should have a positive effect on reducing pipe breaks as observed in the theoretical analysis of Chapter 4.

9.1.3 Evaluation of the use of CPIS in machine learning models for pipe break prediction

Investigation of the impact of CPIS in past research using linear models motivated further research into more advanced analysis. With the theoretical validations of Chapter 4, Chapter 5 performed such analysis using random forest models. The intention of this analysis was to better evaluate the impact of CPIS on predicting pipe breaks, and to provide a methodology that can guide industry professionals into understanding the relevance of mean pressure and dynamic pressure (DP) for pipe break analysis.

The modelling approach with random forest showed that the use of pressure related metrics significantly increases pipe break model accuracy. In particular, for the first presented case study, the combined use of mean pressure and dynamic pressure (DP) increased model accuracy from a R-squared of 0.56 (without the use of pressure components) to a R-squared of 0.81, proving the explanatory power of these variables. The results also showed that DP has been effective in capturing the effects of pressure variations. A comparison of the results obtained by a model utilising the DP variable against a model using pressure range, showed that DP has greater fitting capabilities. The combination of DP and pressure range in a single model does not add additional benefits to modelling pipe breaks, suggesting that information of pressure range is practically contained with DP. DP is calculated using pressure cycle counts. Pressure time series of water pressure can contain several small cycles which might cause a computational burden. It was examined that utilising a threshold of 5 m for minimum pressure cycle range provided similar results as using lower thresholds for cycle range, suggesting the use of a threshold in the range of 5 m to obtain appropriate results. Higher thresholds resulted in reduced explanatory power.

In general, in both case studies presented in Chapter 5, the inclusion of CPIS (given by mean pressure and DP), obtained from calibrated models using actual pressure readings increased the models' power performance. The increase in model performance was accompanied by a reduction of the feature importance of typical covariates used in pipe break prediction such as pipe length. This reduction highlights the importance of the pressure variables at better explaining pipe breaks. The results presented in Chapter 5 demonstrate the significance of ongoing monitoring and systematic gathering of high-resolution pressure data as integral elements of a comprehensive asset management strategy for water distribution networks. The presented methodology using random forest has concluded that CPIS as represented by mean pressure and DP is relevant at explaining pipe breaks from observational data. Therefore, further analysis in larger experimental studies should be sought to establish CPIS as an industry standard for pipe operational asset health.

9.1.4 Determination of the optimal sampling rate for reliable cycle counting in water pressure time series used in CPIS calculation

Chapter 6 investigated the trade-off between accuracy and sampling rate for pressure cycle count calculation using high frequency pressure data. Higher frequency sampling rates increase energy consumption, therefore, identifying the minimal sampling rate that produces reliable cycle count results could extend the lifespan of battery-powered pressure loggers. This is critical for the planning and management of continuous high frequency pressure monitoring needed for CPIS analysis.

Results from analysing two datasets of one year pressure monitored at 128 samples per second (S/s) showed that a sampling rate of 10 S/s is an optimal rate for the sole purpose of counting cycles of water pressure. It was estimated that this sampling rate could produce tolerable losses on average of 6 to 11% of pressure cycles greater than 5 m at different range bands. This magnitude aligns well with results of Chapter 5 that indicated that considering pressure cycles of 5 m and above can produce reliable results for pipe break analysis using CPIS. Further downsampling of sampling rates could be done up to 1 S/s but at the cost of losing approximately 15-25% on average per range band for pressure cycles over 5 m. Conclusions from this chapter suggest that if compensatory measures are implemented to account for the loss of cycle counts, even a sampling rate of 1 S/s could be considered adequate for cycle counting. Downsampling pressure monitoring ideally to 10 S/s with a minimum of 1S/s would broaden the range of feasible options for acquiring high-resolution pressure data necessary for CPIS computation, and extending the autonomy of battery powered sensors.

9.1.5 Method for interpolating a dynamic pressure metric used in CPIS calculations for each pipe in a network

Chapter 7 investigated a methodology to interpolate the dynamic pressure (DP) metric from a discrete set of pressure monitoring locations in water distribution networks. The interpolation of DP presents a major challenge for the computation of CPIS, as traditional transient pressure models are complex, somewhat inaccurate for the calibration of pressure reflections, and computationally expensive to simulate pressure transients for the time periods needed for DP calculation. Given these constraints, a methodology that leverages the use of interpolation techniques for spatially distributed data and short period transient simulations was proposed and evaluated.

The results of the analysis utilising the proposed methodology showed that interpolation of DP can be effectively performed utilising interpolation techniques for spatially distributed data, particularly Inverse Distance Weighting (IDW). However, an improved interpolation can be obtained if DP is decomposed into two components: one associated with large pressure cycles, and another of smaller pressure cycles, DP_{high} and DP_{low}, respectively.

Results of the methodology showed that DP_{high} can be effectively interpolated by scaling the outputs of a transient model to the magnitudes of measured data. In addition, comparable results can be obtained for DP_{high} by using spatial interpolation techniques, such as IDW. DP_{low} can only be interpolated using spatial interpolation techniques, and the accuracy of the results are more dependant on the density and distribution of sensors than DP_{high}. Overall, Chapter 7 demonstrated that spatial interpolation methods can reliably estimate DP at unmonitored locations, provided that a water distribution network has a sufficient number of sensors to capture its key pressure dynamics.

9.1.6 Method to localise sources of transients in networks with uncertain connectivity

The final chapter of the main body of this thesis, Chapter 8 investigated the localisation of sources of pressure transients in water distribution networks. The localisation of pressure transient sources is a critical technical requirement for the interpolation of CPIS as it was presented in Chapter 7.

The literature for pressure transient source localisation has highlighted a method that utilises pressure wave arrival times. However, a critical concern with this method is that uncertainty in network connectivity, as it is common in water networks, can lead to large errors in source localisation. To address this issue, the method investigated in Chapter 8 proposed the use of the k-shortest path routing problem in combination with mixed-integer programming to search a combinatorial space of multiple potential solutions for the source of pressure transients that match readings from measured data. The conclusions of this chapter backed up by an extensive set of simulated data and a real case study showed that the proposed method can accurately determine the source of pressure transients even in mismatched conditions between hydraulic models and actual network scenarios. These results not only prove the use of this method to localise sources of pressure transients, but also as a method to indicate potential problems in network connectivity in hydraulic models, therefore, providing a versatile tool for the asset management of water networks, with focus on CPIS analysis.

9.2 Limitations and future work

Methods to evaluate the impact of pressure transients through the cumulative pressure-induced stress concept have been thoroughly evaluated in this thesis. These evaluation methods covered three main foundational areas of motivation, analysis, and methods for the automated calculation of CPIS. The research presented through these areas has

motivated the following directions for future work:

- **Assessment of CPIS on experimental programs**. Clearly, the work produced in this thesis aimed at enabling the CPIS concept to be advanced to a level where it can be tested in experimental programs. The foundational analysis provided supports the influence of pressure transients on pipe breaks. The developed random forest models also show that CPIS contributes to the explanation of pipe breaks in pipe breaks prediction models. Therefore, ultimate confirmation of these results should be sought by testing CPIS in controlled experiments. To prepare these experiments, methods for the interpolation of CPIS have also been provided. What remains for a large experiment to take place is to combine the findings of this thesis and the use of design of experiments theory to produce a reliable experimental program design. The experimental program should aim at testing experimental and control groups of DMAs, with and without the application of pressure management for calm water networks. The monitoring of these control and experimental groups over time, combined with modelling approaches would ensure the capabilities and limitations of CPIS as a metric for pipe asset health.

- **Formulation of dynamic pressure**. Dynamic Pressure (DP) was tested in this thesis using the original formulation presented in Rezaei (2017). The findings of Chapter 5 confirmed that this metric was reliable for assessing the impacts of pressure transients in the presented case study. However, on the basis of fatigue S-N curves, and fatigue crack propagation, it is sensible to pose that the current metric formulation might be hindering the impacts of large pressure transients with the simple linear multiplication of pressure cycles and their range. The DP data used in the analysis of Chapter 5, was already pre-calculated from a past research, not allowing to evaluate options for its formulation. Therefore, additional formulations of DP should be investigated, for example, by the addition of an exponent factor for the magnitude of pressure cycles, and comparing the results to models that just take in raw input variables from cycle counting.

- **Detailed fatigue assessment for other pipe materials**. This thesis focused almost exclusively on the assessment of cast iron pipes, which was influenced by their large extend in UK water distribution networks. Cast iron pipes are the oldest group of pipes still in operation and most affected by deterioration. However, in the years to come, deterioration of newer plastic pipes could carry additional problems for asset management for which water engineers need to be prepared. These pipes are also characterised by much lower fracture toughness, and a different fatigue crack propagation behaviour which should be investigated in more detail in future research, in particular to assess the impacts of background transients.

- **Location-dependent sampling rate** This thesis investigated the sampling rate requirements for CPIS calculation by evaluating the loss of cycle count at different sampling resolutions. The analysis was conducted using data from a sensor positioned near a transient pressure source. However, since pressure waves change shape as they propagate, dynamic, location-dependent sampling strategies could be explored, adjusting the sampling rate based on the sensor's proximity to transient pressure sources. This approach may help optimize battery consumption by allowing certain sensors to adjust their sampling rates depending on expected pressure waveform characteristics.

- **Refinement of pressure transient source localisation**. This thesis has proposed a novel transient source localisation method. Whilst this method was tested on extensive simulated data and a real operational case study, there is still room for development. In particular, the proposed method formulated a weighting function to address uncertainty in pressure wave arrival times. Future research should explore the inclusion of probabilistic estimates of pipe wave speeds in the problem formulation in order to obtain robust estimations for the solution of pressure transient source localisation. In addition, the application of the method should be investigated in the presence of multiple simultaneous pressure transient sources.

- **Sensor placement for robust interpolation and localisation of transient**

pressure. This thesis investigated the interpolation and localisation of pressure transients for CPIS assessment. The sensors used for the analysis were already in place in the different case studies from previous research. The optimal placement of sensors is usually guided by the expertise of water operators and in many cases constrained by topological and technical limitations. Research on optimal sensor placement has been studied for a variety of problems to optimise, for example, leak localisation, network calibration, and contaminant detection, among others. Placement of sensor for the particular purpose of estimating CPIS has not been investigated in detail. Such investigation would better inform the design of experiments presented in the first bullet point of this section. Optimal sensor placement for transient source localisation has been investigated in Zecchin et al. (2022). However, the method for pressure transient source localisation slightly differs from the one used in this thesis. In addition, investigation of robust optimisation on the formulations for sensor placements is encouraged for more reliable solutions.

Bibliography

Agrawal, A., Verschueren, R., Diamond, S. & Boyd, S. (2018) A rewriting system for convex optimization problems. *Journal of Control and Decision*. 5(1), 42–60. doi: 10.1080/23307706.2017.1397554.

Andaur Navarro, C. L., Damen, J. A., Takada, T., Nijman, S. W., Dhiman, P., Ma, J., Collins, G. S., Bajpai, R., Riley, R. D., Moons, K. G. & Hooft, L. (2021) Risk of bias in studies on prediction models developed using supervised machine learning techniques: Systematic review. *The BMJ*. 375. doi: 10.1136/bmj.n2281.

Anderson, T. L. (2017) *Fracture Mechanics: Fundamentals and Applications*. fourth edition edn. Boca Raton: CRC Press, Taylor & Francis Group.

Andreou, S. A., Marks, D. H. & Clark, R. M. (1987) A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. *Advances in Water Resources*. 10(1), 2–10. doi: 10.1016/0309-1708(87)90002-9.

Anzai, H. & Endo, T. (1979) On-site indication of fatigue damage under complex loading. *International Journal of Fatigue*. 1(1), 49–57. doi: 10.1016/0142-1123(79)90044-6.

ASTM (2017) ASTM E1049-85. Standard Practices for Cycle Counting in Fatigue Analysis. https://www.astm.org/e1049-85r17.html

ASTM (2019) ASTM 247-19. Standard Test Method for Evaluating the Microstructure of Graphite in Iron Castings. https://www.astm.org/a0247-19.html

Baicchi, P., Collini, L. & Riva, E. (2007) A methodology for the fatigue design of

notched castings in gray cast iron. *Engineering Fracture Mechanics*. 74(4), 539–548. doi: 10.1016/j.engfracmech.2006.04.018.

Baker, T. J. (1978) The fracture resistance of flake graphite cast iron. *International Journal of Materials in Engineering Applications*. 1(1), 13–18. doi: 10.1016/0141-5530(78)90003-1.

Barfuss, S. L. (2023) Water Main Break Rates in the USA and Canada: A Comprehensive Study. Technical report. Utah Water Research Laboratory.

Barton, N. A., Farewell, T. S. & Hallett, S. H. (2020) Using generalized additive models to investigate the environmental effects on pipe failure in clean water networks. *npj Clean Water*. 3(1). doi: 10.1038/s41545-020-0077-3.

Barton, N. A., Farewell, T. S., Hallett, S. H. & Acland, T. F. (2019) Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Research*. 164. doi: 10.1016/j.watres.2019.114926.

Barudžija, U., Ivšinović, J. & Malvić, T. (2024) Selection of the Value of the Power Distance Exponent for Mapping with the Inverse Distance Weighting Method—Application in Subsurface Porosity Mapping, Northern Croatia Neogene. *Geosciences (Switzerland)*. 14(6). doi: 10.3390/geosciences14060155.

Bergant, A., Simpson, A. R. & Vítkovský, J. (2001) Developments in unsteady pipe flow friction modelling. *Journal of Hydraulic Research*. 39(3), 249–257. doi: 10.1080/00221680109499828.

Boulos, P. F., Karney, B. W., Wood, D. J. & Lingireddy, S. (2005) Hydraulic transient guidelines for protecting water distribution systems. *Journal - American Water Works Association*. 97(5), 111–124. doi: 10.1002/j.1551-8833.2005.tb10892.x.

Boxall, J. B., O'Hagan, A., Pooladsaz, S., Saul, A. J. & Unwin, D. M. (2007) Estimation of burst rates in water distribution mains. *Proceedings of the Institution of Civil Engineers: Water Management*. 160(2), 73–82. doi: 10.1680/wama.2007.160.2.73.

Breiman, L. (2001) Random Forests. *Machine Learning*. 45, 5–32. doi: 10.1023/A:1010933404324.

Breiman, L. & Cutler, A. (2004) Random Forests. https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm#ooberr

British Standards Institute (1917) BS 78-1917. Specification for cast iron pipes and special castings for water, gas and sewage.

British Standards Institute (1958) BS 1211:1958. Specification for centrifugally cast (spun) iron pressure pipes for water, gas and sewage.

Bulloch, J. H. (1995) Near threshold fatigue behaviour of flake graphite cast irons microstructures. *Theoretical and Applied Fracture Mechanics*. 24(1), 65–78. doi: 10.1016/0167-8442(95)00032-A.

Burhan, I. & Kim, H. S. (2018) S-N curve models for composite materials characterisation: An evaluative review. *Journal of Composites Science*. 2(3). doi: 10.3390/jcs2030038.

Cast Iron Pipe Research Association (1952) *Handbook of Cast Iron Pipe Water, Gas, Sewerage and Industrial Service Second Edition*. Cast Iron Pipe Research Association.

Changklom, J. & Stoianov, I. (2019) Redundant flow estimation methods for robust hydraulic control in water supply networks. *Journal of Hydroinformatics*. 21(4), 571–592. doi: 10.2166/hydro.2019.099.

Cheng, Y.-W. & Broz, J. J. (1986) Cycle-counting methods for fatigue analysis with random load histories : A Fortran User's Guide. Technical report. National Bureau of Standards. Gaithersburg, MD. https://nvlpubs.nist.gov/nistpubs/Legacy/IR/nbsir86-3055.pdf

Chester, R., Jerosch-Herold, C., Lewis, J. & Shepstone, L. (2018) Psychological factors are associated with the outcome of physiotherapy for people with shoulder pain: A multicentre longitudinal cohort study. *British Journal of Sports Medicine*. 52(4), 269–275. doi: 10.1136/bjsports-2016-096084.

Chilès, J. P. & Desassis, N. (2018) Fifty years of kriging. *in* Handbook of Mathematical Geosciences: Fifty Years of IAMG. Springer International Publishing. pp. 589–612.

Choi, K. & Chong, K. (2022) Modified Inverse DistanceWeighting Interpolation for Particulate Matter Estimation and Mapping. *Atmosphere*. 13(5). doi: 10.3390/atmos13050846.

Conlin, R. M. (1991) The Application of Fracture Mechanics to Grey Cast Iron Pipework. PhD thesis. Imperial College London (University of London).

Covas, D., Stoianov, I., Ramos, H., Graham, N., Maksimovic, C. & Butler, D. (2004) Water hammer in pressurized polyethylene pipes: conceptual model and experimental analysis. *Urban Water Journal*. 1(2), 177–197. doi: 10.1080/15730620412331289977.

Cullin, M. J., Petersen, T. H. & Paris, A. (2015) Corrosion Fatigue Failure of a Gray Cast Iron Water Main. *Journal of Pipeline Systems Engineering and Practice*. 6(2), 05014003. doi: 10.1061/(asce)ps.1949-1204.0000188.

Dawood, T., Elwakil, E., Novoa, H. M. & Delgado, J. F. G. (2020) Artificial intelligence for the modeling of water pipes deterioration mechanisms. *Automation in Construction*. 120. doi: 10.1016/j.autcon.2020.103398.

Debnath, S., Ali, I. M., Dhar, A. S. & Thodi, P. (2021) Material properties for fracture mechanics based strength assessment of cast iron water mains. *Canadian Journal of Civil Engineering*. 48(1), 62–74. doi: 10.1139/cjce-2019-0229.

Delnaz, A., Nasiri, F. & Li, S. S. (2023) Asset management analytics for urban water mains: a literature review. *Environmental Systems Research*. 12(1). doi: 10.1186/s40068-023-00287-7.

Diamond, S. & Boyd, S. (2016) CVXPY: A Python-Embedded Modeling Language for Convex Optimization. *Journal of Machine Learning Research*. 17, 1–5. http://jmlr.org/papers/v17/15-408.html

Dijkstra, E. W. (1959) A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik*. 1, 269–271. doi: 10.1007/BF01386390.

Dowling, N. E. (1972) Fatigue Failure Predictions for Complicated Stress-Strain Histories. *Journal of Materials, JMLSA*. 7(1), 71–87.

Economic Insight (2022) Options for a Sustainable Approach to Asset Maintenance and Replacement. A Report for Water UK. Technical report. Water UK. https://www.water.org.uk/sites/default/files/wp/2022/06/Options-for-a-Sustainable-Approach-to-Asset-Maintenance-and-Replacement-June-2022.pdf

Elber, W. (1970) Fatigue crack closure under cyclic tension. *Engineering Fracture Mechanics*. 2(1), 37–44. doi: 10.1016/0013-7944(70)90028-7.

Ellis, J. (2008) *Pressure transients in water engineering*. Thomas Telford Publishing. https://www.icevirtuallibrary.com/doi/abs/10.1680/ptiwe.35928

Environmental Protection Agency (2023) Drinking Water Infrastructure Needs Survey and Assessment. 7 th Report to Congress. Technical report. Environmental Protection Agency. https://www.epa.gov/system/files/documents/2023-09/Seventh%20DWINSA_September2023_Final.pdf

Fan, X., Wang, X., Zhang, X. & ASCE Xiong (Bill) Yu, P. E. (2022) Machine learning based water pipe failure prediction: The effects of engineering, geology, climate and socio-economic factors. *Reliability Engineering and System Safety*. 219. doi: 10.1016/j.ress.2021.108185.

Fantozzi, M., Calza, F. & Lambert, A. (2009) Experience and results achieved in introducing District Metered Areas (DMA) and Pressure Management Areas (PMA) at Enia utility (Italy). *in* Proceedings of the 5th IWA Water Loss Reduction Specialist Conference. pp. 153–160.

Fantozzi, M., Lamber, A., Charalambous, B., Pothof, I. & Kovac, J. (2013) Training_coaching utilities in practical implementation of latest developments in Pressure

Management. *in* 7th IWA International Conference on Efficient Use and Management of Water (Efficient 2013).

Fatemi, A. & Vangt, L. (1998) Cumulative fatigue damage and life prediction theories: a survey of the state of the art for homogeneous materials. *International Journal of Fatigue*. 20(1), 9–34. doi: 10.1016/S0142-1123(97)00081-9.

Fu, G., Zhang, C., Deo, R., Rathnayaka, S., Shannon, B. & Kodikara, J. (2020) A model of stress concentration factors for external corrosion patches on large-diameter underground cast iron pipes. *Sustainable and Resilient Infrastructure*. . doi: 10.1080/23789689.2020.1828022.

Gangl, G., Fuchs-Hanusch, D., Stadlober, E. & Kauch, P. (2007) Analysis of the failure behaviour of drinking water pipelines. *Water Science and Technology: Water Supply*. 7(5-6), 219–225. doi: 10.2166/ws.2007.102.

Garcia Cardona, J., Ortega, A. & Rodriguez-Alvarez, N. (2022) Graph-Based Interpolation for Remote Sensing Data. *in* 2022 30th European Signal Processing Conference (EUSIPCO). European Association for Signal Processing (EUSIPCO). doi: 10.23919/EUSIPCO55093.2022.9909647.

Giustolisi, O., Laucelli, D. & Savic, D. A. (2006) Development of rehabilitation plans for water mains replacement considering risk and cost-benefit assessment. *Civil Engineering and Environmental Systems*. 23(3), 175–190. doi: 10.1080/10286600600789375.

Gringarten, E. & Deutsch, C. V. (2001) Teacher's Aide Variogram Interpretation and Modeling. *Mathematical Geology*. 33(4). doi: 10.1023/A:1011093014141.

Guibert, R., Bayle, A. & Plouraboué, F. (2023) Geolocalization of water-waves origin within water distribution networks using time reversal of first event detection. *Water Research*. 230. doi: 10.1016/j.watres.2022.119538.

Gurobi Optimization LLC (2023) Gurobi Optimizer Reference Manual. https://docs.gurobi.com/projects/optimizer/en/current/index.html

Habibian, A. (1994) Effect of Temperature Changes on Water-Main Breaks. *Journal of Transportation Engineering*. 120(2), 312–321. doi: 10.1061/(ASCE)0733-947X(1994)120:2(312).

Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H. & Bing, G. (2017) Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*. 73, 220–239. doi: 10.1016/j.eswa.2016.12.035.

Hampson, W. J. (2014) Locating Sources of Pressure Transients in Water Distribution Systems. PhD thesis. The University of Sheffield. Sheffield.

Hastie, T., Tibshirani, R. & Friedman, J. (2017) *Springer Series in Statistics*. second edition edn. Springer. http://www.springer.com/series/692

Hornbogen, E. (1985) Fracture toughness and fatigue crack growth of grey cast irons. *Journal of Materials Science*. 20, 3897–3905. doi: 10.1007/BF00552378.

Hoskins, A. J. (2015) Monitoring the Hydraulic Dynamics in Water Distribution Systems. PhD thesis. Imperial College London. London. doi: 10.25560/58335. http://hdl.handle.net/10044/1/58335

Hoskins, A. & Stoianov, I. (2014) InfraSense: A distributed system for the continuous analysis of hydraulic transients. *in* Procedia Engineering. Vol. 70. Elsevier Ltd. pp. 823–832. doi: 10.1016/j.proeng.2014.02.090.

Hoskins, A. & Stoianov, I. (2017) Monitoring fluid dynamics. WO 2017/060737(Patent).

Hussein Farh, H. M., Ben Seghier, M. E. A., Taiwo, R. & Zayed, T. (2023) Analysis and ranking of corrosion causes for water pipelines: a critical review. *npj Clean Water*. 6(1). doi: 10.1038/s41545-023-00275-5.

Irwin, G. R. (1956) Onset of fast crack propagation in high strength steel and aluminum alloys. Technical report. Naval Research Laboratoy. Washington D.C.

James, M. N. & Wenfong, L. (1999) Fatigue crack growth in austempered ductile and grey cast irons-stress ratio effects in air and mine water. *Materials Science and Engineering*. 265, 129–139. doi: 10.1016/S0921-5093(98)01140-X.

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

Jara-Arriagada, C. & Stoianov, I. (2022) High resolution water pressure monitoring for the assessment of fatigue damage in water distribution pipes. *in* Proceedings - 2nd International Joint Conference on Water Distribution System Analysis (WDSA)& Computing and Control in the Water Industry (CCWI). Editorial Universitat Politècnica de València. València. doi: 10.4995/WDSA-CCWI2022.2022.14813. http://ocs.editorial.upv.es/index.php/WDSA-CCWI/WDSA-CCWI2022/paper/view/14813

Jara-Arriagada, C. & Stoianov, I. (2024) Pressure-induced fatigue failures in cast iron water supply pipes. *Engineering Failure Analysis*. 155, 107731. doi: 10.1016/j.engfailanal.2023.107731. https://linkinghub.elsevier.com/retrieve/pii/S1350630723006854

Jara-Arriagada, C., Ulusoy, A.-J. & Stoianov, I. (2024) Localization of Transient Pressure Sources in Water Supply Networks with Connectivity Uncertainty. *Journal of Water Resources Planning and Management*. 150(4). doi: 10.1061/JWRMD5.WRENG-6235. https://ascelibrary.org/doi/10.1061/JWRMD5.WRENG-6235

Kang, J., Park, Y. J., Lee, J., Wang, S. H. & Eom, D. S. (2018) Novel leakage detection by ensemble CNN-SVM and graph-based localization in water distribution systems. *IEEE Transactions on Industrial Electronics*. 65(5), 4279–4289. doi: 10.1109/TIE.2017.2764861.

Karney, B. W. & McInnis, D. (1990) Transient analysis of water distribution systems.

Journal - American Water Works Association. 82(7), 62–70. doi: 10.1002/j.1551-8833.1990.tb06992.x.

Kim, J., Bae, C., Woo, H., Kim, J. & Hong, S. (2007) Assessment of residual tensile strength on cast iron pipes. *in* International Conference on Pipeline Engineering and Construction.

Kleiner, Y. & Rajani, B. (2001) Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water*. 3, 131–150. doi: 10.1016/S1462-0758(01)00033-4.

Konstantinou, C., Jara-Arriagada, C. & Stoianov, I. (2024) Investigating the Impact of Cumulative Pressure-Induced Stress on Machine Learning Models for Pipe Breaks. *Water Resources Management*. 38, 603–619. doi: 10.1007/s11269-023-03687-7. https://link.springer.com/10.1007/s11269-023-03687-7

Konstantinou, C. & Stoianov, I. (2020) A comparative study of statistical and machine learning methods to infer causes of pipe breaks in water supply networks. *Urban Water Journal*. 17(6), 534–548. doi: 10.1080/1573062X.2020.1800758.

Kuha, J. (2004) AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods and Research*. 33(2), 188–229. doi: 10.1177/0049124103262065.

Larrosa, N. O., Akid, R. & Ainsworth, R. A. (2018) Corrosion-fatigue: a review of damage tolerance models. *International Materials Reviews*. 63(5), 283–308. doi: 10.1080/09506608.2017.1375644.

Laven, K., Ge, S., Dignum, M., Zantingh, M., Gill, N., Ponnambalam, K. & Krinas, H. (2024) Pipe Diameter Is a Dominant Factor in Pipe Failure Risk. *in* Pipelines 2024. pp. 489–498.

Li, C. Q. & Mahmoodian, M. (2013) Risk based service life prediction of underground cast iron pipes subjected to corrosion. *Reliability Engineering and System Safety*. 119, 102–108. doi: 10.1016/j.ress.2013.05.013.

Li, J. & Heap, A. D. (2014) Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling and Software*. 53, 173–189. doi: 10.1016/j.envsoft.2013.12.008.

Li, W., Mazumder, R. K. & Li, Y. (2021) Time-Dependent Reliability Analysis of Buried Water Distribution Network: Combined Finite-Element and Probabilistic Approach. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*. 7(4), 04021064. doi: 10.1061/ajrua6.0001178.

Liaw, A. & Wiener, M. (2002) Classification and Regression by randomForest. *The Newsletter of the R Project News*. 2(3), 18–22. http://www.stat.berkeley.edu/

Liu, Z., Sadiq, R., Rajani, . B. & Najjaran, H. (2010) Exploring the Relationship between Soil Properties and Deterioration of Metallic Pipes Using Predictive Data Mining Methods. *Journal of Computing in Civil Engineering*. 24(3), 289–301. doi: 10.1061/ASCECP.1943-5487.0000032.

Liyanage, K. T. H. & Dhar, A. S. (2017) Effects of Corrosion Pits on Wall Stresses in Cast-Iron Water Mains. *Journal of Pipeline Systems Engineering and Practice*. 8(4), 04017023. doi: 10.1061/(asce)ps.1949-1204.0000286.

Lloyd, C. D. (2005) Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. *Journal of Hydrology*. 308(1-4), 128–150. doi: 10.1016/j.jhydrol.2004.10.026.

Lu, G. Y. & Wong, D. W. (2008) An adaptive inverse-distance weighting spatial interpolation technique. *Computers and Geosciences*. 34(9), 1044–1055. doi: 10.1016/j.cageo.2007.07.010.

Makar, J. (1999) Failure analysis for grey cast iron water pipes. *in* AWWA Distribution System Symposium.

Makar, J., Desnoyers, R. . & Mcdonald, S. . (2001) Failure modes and mechanisms in gray cast iron pipe NRCC-44218. *in* Underground Infrastructure Research: Municipal, Industrial and Environmental Applications. Kitchener, Ontario. pp. 1–10.

Makar, J. M. (2000) A preliminary analysis of failures in grey cast iron water pipes. *Engineering Failure Analysis*. 7(1), 43–53. doi: 10.1016/S1350-6307(99)00005-9.

Makar, J. M. & Mcdonald, S. E. (2007) Mechanical Behavior of Spun-Cast Gray Iron Pipe. *Journal of Materials in Civil Engineering*. 19(10), 826–833. doi: 10.1061/(ASCE)0899-1561(2007)19:10(826).

Makar, J. M. & Rajani, B. (2000) Gray Cast-Iron Water Pipe Metallurgy. *Journal of Materials in Civil Engineering*. 12(3), 245–253. doi: 10.1061/(ASCE)0899-1561(2000)12:3(245).

Marlow, D., Gould, S., Beale, D. & Lane, B. (2015) Rehabilitation of small-diameter cast-iron pipe: US, UK, and Australian perspectives. *Journal - American Water Works Association*. 107(1), E12–E21. doi: 10.5942/jawwa.2015.107.0003.

Marlow, D., Gould, S. & Lane, B. (2015) An expert system for assessing the technical and economic risk of pipe rehabilitation options. *Expert Systems with Applications*. 42(22), 8658–8668. doi: 10.1016/j.eswa.2015.07.020.

Marsili, V., Alvisi, S., Maietta, F., Capponi, C., Meniconi, S., Brunone, B. & Franchini, M. (2023) Extending the Application of Connectivity Metrics for Characterizing the Dynamic Behavior of Water Distribution Networks. *Water Resources Research*. 59(8). doi: 10.1029/2023WR035031.

Marsili, V., Meniconi, S., Alvisi, S., Brunone, B. & Franchini, M. (2021) Experimental analysis of the water consumption effect on the dynamic behaviour of a real pipe network. *Journal of Hydraulic Research*. 59(3), 477–487. doi: 10.1080/00221686.2020.1780506.

Martin, A., Delgado-Aguiñaga, J. A. & Puig, V. (2022) Control of transients in drinking water networks. *Control Engineering Practice*. 119. doi: 10.1016/j.conengprac.2021.104986.

Martínez-Codina, Castillo, M., González-Zeas, D. & Garrote, L. (2016) Pressure as a

predictor of occurrence of pipe breaks in water distribution networks. *Urban Water Journal*. 13(7), 676–686. doi: 10.1080/1573062X.2015.1024687.

Martínez García, D., Lee, . J., Asce, M., Keck, J., Kooy, J., Yang, P. & Wilfley, B. (2020) Pressure-Based Analysis of Water Main Failures in California. *Journal of Water Resources Planning and Management*. 149(9). doi: 10.1061/(ASCE) WR.1943-5452.0001255.

Mazumder, R. K., Salman, A. M., Li, Y. & Yu, X. (2019) Reliability Analysis of Water Distribution Systems Using Physical Probabilistic Pipe Failure Method. *Journal of Water Resources Planning and Management*. 145(2). doi: 10.1061/(asce)wr.1943-5452.0001034.

McInnis, D. (2004) A relative-risk framework for evaluating transient pathogen intrusion in distribution systems. *Urban Water Journal*. 1(2), 113–127. doi: 10.1080/15730620412331290010.

McInnis, D. & Karney, B. (1995) Transients in Distribution Networks: Field Tests and Demand Models. *Journal of Hydraulic Engineering*. 121(3), 218–231.

Milà, C., Mateu, J., Pebesma, E. & Meyer, H. (2022) Nearest neighbour distance matching Leave-One-Out Cross-Validation for map validation. *Methods in Ecology and Evolution*. 13(6), 1304–1316. doi: 10.1111/2041-210X.13851.

Miner, M. A. (1945) Cumulative Damage in Fatigue. *Journal of Applied Mechanics*. 12(3), A159–A164. doi: 10.1115/1.4009458.

Misiunas, D., Lambert, M., Simpson, A. & Olsson, G. (2005) Burst detection and location in water distribution networks. *Water Science and Technology: Water Supply*. 5(3-4), 71–80. http://iwaponline.com/ws/article-pdf/5/3-4/71/417882/71.pdf

Misiunas, D., Vítkovský, J., Olsson, G., Simpson, A., Asce, M. & Lambert, M. (2005) Pipeline Break Detection Using Pressure Transient Monitoring. *Journal of Water Resources Planning and Management*. 131(4), 316–325. doi: 10.1061/ASCE0733-94962005131:4316.

Mohebbi, H., Jesson, D. A., Mulheron, M. J. & Smith, P. A. (2010) The fracture and fatigue properties of cast irons used for trunk mains in the water industry. *Materials Science and Engineering A*. 527(21-22), 5915–5923. doi: 10.1016/j.msea.2010.05.071.

Molnar, C. (2018) iml: An R package for Interpretable Machine Learning. *Journal of Open Source Software*. 3(26), 786. doi: 10.21105/joss.00786.

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

Narang, S. K., Gadde, A. & Ortega, A. (2013) Signal processing techniques for interpolation in graph structured data. *in* 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. doi: 10.1109/ICASSP.2013.6638704.

Newman, J. C. (1976) Fracture Analysis of Surface and Through Cracks in Cylindrical Pressure Vessels. Technical report. Nasa Technical Note.

Ng, A. (2007*a*) CS229 - Machine Learning, Lecture Notes: Regularization and model selection. https://see.stanford.edu/materials/aimlcs229/cs229-notes5.pdf

Ng, A. (2007*b*) CS229 - Machine Learning, Lecture Notes: Supervised learning. https://see.stanford.edu/materials/aimlcs229/cs229-notes1.pdf

Nishiyama, M. & Filion, Y. (2013) Review of statistical water main break prediction models. *Canadian Journal of Civil Engineering*. 40(10), 972–979. doi: 10.1139/cjce-2012-0424.

Ozelkan, E., Chen, G. & Ustundag, B. B. (2016) Spatial estimation of wind speed: a new integrative model using inverse distance weighting and power law. *International Journal of Digital Earth*. 9(8), 733–747. doi: 10.1080/17538947.2015.1127437.

Pearson, D., Fantozzi, M., Soares, D. & Waldron, T. (2005) Pearson - Searching for N2_ How does Pressure Reduction Reduce Burst Frequency. *in* Proc. Intl. Water Assoc. Leakage Conference.

Pecci, F., Abraham, E. & Stoianov, I. (2019) Model Reduction and Outer Approximation for Optimizing the Placement of Control Valves in Complex Water Networks. *Journal of Water Resources Planning and Management*. 145(5). doi: 10.1061/(asce)wr.1943-5452.0001055.

Pecci, F., Parpas, P. & Stoianov, I. (2020) Sequential Convex Optimization for Detecting and Locating Blockages in Water Distribution Networks. *Journal of Water Resources Planning and Management*. 146(8). doi: 10.1061/(ASCE)WR.1943. https://orcid.org/0000-0003-3200-0892.

Pelletier, G., Mailhot, A. & Villeneuve, J.-P. (2003) Modeling Water Pipe Breaks-Three Case Studies. *Journal of Water Resources Planning and Management*. 129(2), 115–123. doi: 10.1061/ASCE0733-94962003129:2115.

Petersen, R. B. & Melchers, R. E. (2012) Long-Term Corrosion of Cast Iron Cement Lined Pipes. *in* Corrosion & Prevention 2012. Melbourne, Australia.

Plouraboué, F. (2024) Review on water-hammer waves mechanical and theoretical foundations. *European Journal of Mechanics, B/Fluids*. 108, 237–271. doi: 10.1016/j.euromechflu.2024.08.001.

Rajani, B. & Kleiner, Y. (2001) Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water 3*. pp. 151–164. doi: 10.1016/S1462-0758(01)00032-2.

Rajani, B., Kleiner, Y. & Sink, J. E. (2012) Exploration of the relationship between water main breaks and temperature covariates. *Urban Water Journal*. 9(2), 67–84. doi: 10.1080/1573062X.2011.630093.

Rajani, B., Lewandowski, J. & Margevicius, A. (2012) Failure analysis of cast iron trunk main in Cleveland, Ohio. *Journal of Failure Analysis and Prevention*. 12(3), 217–236. doi: 10.1007/s11668-012-9567-z.

Rajani, B., Makar, J., Rajani, B. & Makar, J. (2000) A methodology to estimate remaining service life of grey cast iron water mains. *Canadian Journal of Civil Engineering*. 27, 1259–1272.

Ramalingam, D., Lingireddy, S. & Wood, D. J. (2009) Using the WCM for transient modeling of water distribution networks. *Journal - American Water Works Association*. 101(2). doi: 10.1002/j.1551-8833.2009.tb09841.x.

Rathnayaka, S., Shannon, B., Rajeev, P. & Kodikara, J. (2015) Monitoring of pressure transients in water supply networks. *Water Resources Management*. 30(2), 471–485. doi: 10.1007/s11269-015-1172-y.

Rathnayaka, S., Shannon, B., Zhang, C. & Kodikara, J. (2017) Introduction of the leak-before-break (LBB) concept for cast iron water pipes on the basis of laboratory experiments. *Urban Water Journal*. 14(8), 820–828. doi: 10.1080/1573062X.2016.1274768.

ReliaSoft Corporation (2024) Miner's Rule and Cumulative Damage Models. https://help.reliasoft.com/articles/content/hotwire/issue116/hottopics116.htm

Rezaei, H. (2017) Impact of pressure fluctuations on pipe failures in water distribution networks. PhD thesis. Imperial College London.

Rezaei, H., Ryan, B. & Stoianov, I. (2015) Pipe failure analysis and impact of dynamic hydraulic conditions in water supply networks. *Procedia Engineering*. 119(1), 253–262. doi: 10.1016/j.proeng.2015.08.883.

Riaño-Briceño, G., Hodges, B. R. & Sela, L. (2022) PTSNet: A Parallel Transient Simulator for Water Transport Networks based on vectorization and distributed computing. *Environmental Modelling and Software*. 158. doi: 10.1016/j.envsoft.2022.105554.

Riaño-Briceño, G., Sela, L. & Hodges, B. R. (2021) Distributed and vectorized method of characteristics for fast transient simulations in water distribution systems. *Computer-Aided Civil and Infrastructure Engineering*. 37(2), 163–184. doi: 10.1111/mice.12709.

Rifaai, T. M., Abokifa, A. A. & Sela, L. (2022) Integrated approach for pipe failure prediction and condition scoring in water infrastructure systems. *Reliability Engineering and System Safety*. 220. doi: 10.1016/j.ress.2021.108271.

Robert, D. J., Chan, D., Rajeev, P. & Kodikara, J. (2022) Effects of operational loads on buried water pipes using field tests. *Tunnelling and Underground Space Technology*. 124. doi: 10.1016/j.tust.2022.104463.

Robinson, T. P. & Metternicht, G. (2003) A comparison of inverse distance weighting and ordinary kriging for characterising within-paddock spatial variability of soil properties in Western Australia. *Cartography*. 32(1), 11–24. doi: 10.1080/00690805.2003.9714231.

Romaniv, O. N., Tkach, A. N., Yus'kiv, T. Y., Ovchinnikov, V. I. & Sharkov, V. A. (1990) Growth of a fatigue crack in cast irons. II. Gray and malleable cast irons. *Soviet materials science : a transl. of Fiziko-khimicheskaya mekhanika materialov / Academy of Sciences of the Ukrainian SSR*. 26(3), 274–281. doi: 10.1007/BF00727357.

Romanoff, M. (1957) Underground Corrosion. Technical report. National Bureau of Standards Circular 579.

Røstum, J. (2000) Statistical modelling of pipe failures in water networks. PhD thesis. Norwegian University of Science and Technology. Trondheim. http://hdl.handle.net/11250/242082

Rychlik, I. (1993) Note on cycle counts in irregular loads. *Fatigue Fruct. Engng Muter. Struct.* 16(4), 377–390. doi: 10.1111/j.1460-2695.1993.tb00094.x.

Sadler, J. M., Goodall, J. L., Morsy, M. M. & Spencer, K. (2018) Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. *Journal of Hydrology*. 559, 43–55. doi: 10.1016/j.jhydrol.2018.01.044.

Saito, T. & Rehmsmeier, M. (2015) The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*. 10(3). doi: 10.1371/journal.pone.0118432.

Scheidegger, A., Leitão, J. P. & Scholten, L. (2015) Statistical failure models for water distribution pipes - A review from a unified perspective. *Water Research*. 83, 237–247. doi: 10.1016/j.watres.2015.06.027.

Shepard, D. (1968) A two-dimensional interpolation for irregularly-spaced data function. *in* ACM '68: Proceedings of the 1968 23rd ACM national conference. pp. 517–524. doi: 10.1145/800186.810616.

Shi, Z., Osher, S. & Zhu, W. (2017) Weighted Nonlocal Laplacian on Interpolation from Sparse Data. *Journal of Scientific Computing*. 73(2-3), 1164–1177. doi: 10.1007/s10915-017-0421-z.

Shiode, N. & Shiode, S. (2011) Street-level Spatial Interpolation Using Network-based IDW and Ordinary Kriging. *Transactions in GIS*. 15(4), 457–477. doi: 10.1111/j.1467-9671.2011.01278.x.

Silhavy, R., Silhavy, P. & Prokopova, Z. (2017) Analysis and selection of a regression model for the Use Case Points method using a stepwise approach. *Journal of Systems and Software*. 125, 1–14. doi: 10.1016/j.jss.2016.11.029.

Silva, N. A., Buiatti, C. M., Cruz, S. L. & F R Pereira, O. A. (1996) Pressure Wave Behaviour and Leak Detection in Pipelines. *Computers chem. Engng*. 20, 491–496.

Soltani Asadi, Z. & Melchers, R. E. (2017*a*) Extreme value statistics for pitting corrosion of old underground cast iron pipes. *Reliability Engineering and System Safety*. 162, 64–71. doi: 10.1016/j.ress.2017.01.019.

Soltani Asadi, Z. & Melchers, R. E. (2017*b*) Pitting corrosion of older underground cast iron pipes. *Corrosion Engineering Science and Technology*. 52(6), 459–469. doi: 10.1080/1478422X.2017.1330025.

Soltani Asadi, Z. & Melchers, R. E. (2018) Clustering of corrosion pit depths for buried cast iron pipes. *Corrosion Science*. 140, 92–98. doi: 10.1016/j.corsci.2018.06.015.

Speidel, M. O. (1981) Bruchzähigkeit und Ermüdungs-Rißwachstum von Gußeisen. *Materialwissenschaft und Werkstofftechnik*. 12(11), 387–402. doi: 10.1002/mawe.19810121106.

Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H. B. & Whittle, A. J. (2013) Wavelet-based burst event detection and localization in water distribution systems. *Journal of Signal Processing Systems*. 72(1), 1–16. doi: 10.1007/s11265-012-0690-6.

Srirangarajan, S., Lim, H. B., Allen, M. & Preis, A. (2010) Water Main Burst Event Detection and Localization. *in* Water Distribution System Analysis 2010. pp. 1324–1335.

Srirangarajan, S. & Pesch, D. (2013) Source localization using graph-based optimization technique. *in* IEEE Wireless Communications and Networking Conference, WCNC. pp. 1127–1132. doi: 10.1109/WCNC.2013.6554722.

Starczewska, D. (2016) Pressure transients in water distribution networks: understanding their contribution to pipe repairs. PhD thesis. The University of Sheffield. Sheffield.

Starczewska, D., Collins, R. & Boxall, J. (2014) Transient behavior in complex distribution network: A case study. *in* Procedia Engineering. Vol. 70. Elsevier Ltd. pp. 1582–1591. doi: 10.1016/j.proeng.2014.02.175.

Starczewska, D., Collins, R. & Boxall, J. (2015) Occurrence of transients in water distribution networks. *in* Procedia Engineering. Vol. 119. Elsevier Ltd. pp. 1473–1482. doi: 10.1016/j.proeng.2016.01.001.

Stoianov, I. & Hoskins, A. (2017) Cumulative Pressure Induced Stress for Characterising Pressure Transients and Variability in Water Supply Networks. *in* The University of Sheffield, ed., CCWI 2017 - Computing and Control for the Water Industry. doi: https://doi.org/10.15131/shef.data.5364700.v1.

Stoianov, I., Maksimovic, C. & Graham, N. (2003) Field Validation of the Application of Hydraulic Transients for Leak Detection in Transmission Pipelines. *in* Advances in Water Supply Management. pp. 86–97.

Stoianov, I., Nachman, L., Madden, S. & Tokmouline, T. (2007) PIPENET: A Wireless Sensor Network for Pipeline Monitoring. *in* IPSN '07: Proceedings of the 6th international conference on Information processing in sensor networks. Association for Computing Machinery. pp. 264–273. doi: 10.1145/1236360.1236396.

Taiwo, R., Ben Seghier, M. E. A. & Zayed, T. (2023) Toward Sustainable Water Infrastructure: The State-Of-The-Art for Modeling the Failure Probability of Water Pipes. *Water Resources Research*. 59(4). doi: 10.1029/2022WR033256.

Taiwo, R., Shaban, I. A. & Zayed, T. (2023) Development of sustainable water infrastructure: A proper understanding of water pipe failure. *Journal of Cleaner Production*. 398. doi: 10.1016/j.jclepro.2023.136653.

Thames Water (2023) Cost Adjustment Claim Mains Replacement. Technical report. Thames Water. https://www.ofwat.gov.uk/wp-content/uploads/2023/08/TMS_mains-replacement.pdf

The MathWorks Inc (2024) stepwiseglm. https://www.mathworks.com/help/stats/stepwiseglm.html

Thomas, R. J. & Peethamparan, S. (2017) Stepwise regression modeling for compressive strength of alkali-activated concrete. *Construction and Building Materials*. 141, 315–324. doi: 10.1016/j.conbuildmat.2017.03.006.

Tornyeviadzi, H. M. & Seidu, R. (2023) Leakage detection in water distribution networks via 1D CNN deep autoencoder for multivariate SCADA data. *Engineering Applications of Artificial Intelligence*. 122. doi: 10.1016/j.engappai.2023.106062.

Twyman, J. (2016) Wave Speed Calculation For Water Hammer Analysis. *Obras y Proyectos*. 20, 86–92.

UKWIR (2022) 22/WM/04/14 - The Occurrence and Causes of Pressure Transients in Distribution Networks. Technical report. UK Water Industry Research.

Ulusoy, A. J., Pecci, F. & Stoianov, I. (2022) Bi-objective design-for-control of water distribution networks with global bounds. *Optimization and Engineering*. 23(1), 527–577. doi: 10.1007/s11081-021-09598-z.

Vítkovský, J. P., Lambert, M. F., Simpson, A. R. & Bergant, A. (2000) Advances in Unsteady Friction Modelling in Transient Pipe Flow. *in* The 8th International Conference on Pressure Surges. The Hague.

Vrieze, S. I. (2012) Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*. 17(2), 228–243. doi: 10.1037/a0027127.

Walker, K. (1970) The Effect of Stress Ratio During Crack Propagation and Fatigue for 2024-T3 and 7075-T6 Aluminum. *in* Effects of Environment and Complex Load History on Fatigue Life, ASTM STP 462. American Society for Testing and Materials. pp. 1–14. doi: 10.1520/STP32032S.

Wang, W., Zhou, A., Fu, G., Li, C. Q., Robert, D. & Mahmoodian, M. (2017) Evaluation of stress intensity factor for cast iron pipes with sharp corrosion pits. *Engineering Failure Analysis*. 81, 254–269. doi: 10.1016/j.engfailanal.2017.06.026.

Wasim, M., Robert, D. & Li, C. Q. (2021) Corrosion effects on the granular nano mechanical properties of buried cast iron. *Engineering Failure Analysis*. 119. doi: 10.1016/j.engfailanal.2020.104970.

Whitley, D. (1994) A genetic algorithm tutorial. *Statistics and Computing*. 4, 65–85.

Winkler, D., Haltmeier, M., Kleidorfer, M., Rauch, W. & Tscheikner-Gratl, F. (2018) Pipe failure modelling for water distribution networks using boosted decision trees. *Structure and Infrastructure Engineering*. 14(10), 1402–1411. doi: 10.1080/15732479.2018.1443145.

Wols, B. A. & Van Thienen, P. (2014) Impact of weather conditions on pipe failure: A statistical analysis. *Journal of Water Supply: Research and Technology - AQUA*. 63(3), 212–223. doi: 10.2166/aqua.2013.088.

Wong, L., Rathnayaka, S., Chiu, W. K. & Kodikara, J. (2017) Fatigue Damage Monitoring of a Cast Iron Pipeline Using Distributed Optical Fibre Sensors. *in* Procedia Engineering. Vol. 188. Elsevier Ltd. pp. 293–300. doi: 10.1016/j.proeng.2017.04.487.

Wright, R., Abraham, E., Parpas, P. & Stoianov, I. (2015) Control of water distribution networks with dynamic DMA topology using strictly feasible sequential convex programming. *Water Resources Research*. 51(12), 9925–9941. doi: 10.1002/2015WR017466.

Xing, L. & Sela, L. (2020) Transient simulations in water distribution networks: TSNet python package. *Advances in Engineering Software*. 149. doi: 10.1016/j.advengsoft.2020.102884.

Yen, J. Y. (1971) Finding the K Shortest Loopless Paths in a Network. *Management Science*. 17(11), 712–716. doi: 10.1287/mnsc.17.11.712. https://about.jstor.org/terms

Zecchin, A. C., Do, N., Gong, J., Leonard, M., Lambert, M. F. & Stephens, M. L. (2022) Optimal Pipe Network Sensor Layout Design for Hydraulic Transient Event Detection and Localization. *Journal of Water Resources Planning and Management*. 148(8). doi: 10.1061/(asce)wr.1943-5452.0001536.

Zhao, Q. & Hastie, T. (2021) Causal Interpretations of Black-Box Models. *Journal of Business and Economic Statistics*. 39(1), 272–281. doi: 10.1080/07350015.2019.1624293.

Zhou, L., Li, Y., Zhao, Y., Ou, C. & Zhao, Y. (2021) An accurate and efficient scheme involving unsteady friction for transient pipe flow. *Journal of Hydroinformatics*. 23(4), 879–896. doi: 10.2166/hydro.2021.160.

Appendix A

Transient model calibration

This appendix presents the process of transient model calibration used to calibrate the transient model used for the calculation of DP_{model} in Chapter 7.

Calibrating the transient model to simulate the large cyclic loadings observed in the recorded water pressure time series first requires identifying the source of these substantial pressure fluctuations. Such sources are generally limited to automated valves, pumps, or large industrial users. Detailed methodologies for localising pressure transient sources within water supply networks have been discussed in the Literature Review (Chapter 2) and Chapter 8.

Once the transient source is identified, a transient model employing the method of characteristics with steady-state friction factors is adjusted (Xing & Sela, 2020) so that its simulated pressure wave amplitudes align with the maximum observed ranges at selected sensor locations. Numerous studies have demonstrated that method-of-characteristics-based models effectively capture the initial, steep rise of pressure transients but often fail to accurately represent their subsequent attenuation and dissipation, largely due to complex wave dissipation mechanisms and other uncertainties.

For the analysis of Chapter 7, a transient model was calibrated by comparing the model results with data from the second sensor closest to the source in the case study. The

second closes sensor to the source was chosen for the smoothness of its recorded pressure wave profile. This simplification allowed for a clearer assessment of the model's ability to reproduce the primary characteristics of the measured transient events.

A.1 Genetic algorithm optimisation of transient parameters

The calibration procedure described in this appendix employed a genetic algorithm (Whitley, 1994) to adjust the transient model parameters. For this purpose, a burst event was selected as a proxy for the transient generated by the large industrial user, given that the waveforms produced by significant bursts are similar in shape. The variables to be calibrated included the burst coefficient, which quantifies the severity of the burst, and the time of maximum burst development (Xing & Sela, 2020).

Starting from an initial parameter guess, the genetic algorithm iteratively modified these parameters, seeking to minimise the discrepancy between the simulated transient pressure wave and the measured data. The objective function was defined as the mean squared error (MSE) between the measured and simulated pressure profiles from the start of the transient until its peak. The genetic algorithm was favoured over alternative optimisation strategies due to the substantial computational time required for each transient simulation (approximately 30 minutes) and its demonstrated capability for rapidly approximating a satisfactory solution.

A.2 Post-processing of transient model results

After calibrating the transient model to align with the observed maximum cycle range, a final post-processing step was applied to reduce the intensity of the simulated wave reflections. Pressure transient simulations that rely on the method of characteristics and

steady friction factors often face difficulties in accurately reproducing wave reflections beyond the primary rising edge, due to a range of factors and uncertainties such as variable demand, entrapped air, pipe material properties, and elasticity (Bergant et al., 2001; Zhou et al., 2021). Although various strategies have been proposed to address these shortcomings, many require more complex modelling approaches.

Considering the relatively uniform pipe materials in the network, a heuristic post-processing adjustment was employed to approximate these unaccounted-for dissipation processes. Specifically, stretching and attenuation factors were introduced to emulate the attenuation patterns observed in the measured data. The stretching factor was assumed to follow:

$$k = 1 + \frac{x}{20\pi} \tag{A.1}$$

where x represents the time step in the time series, and k is the factor to stretch this time step.

the attenuation factor is an exponentially decaying function of the form:

$$d = \exp\left(-\lambda y\right) \tag{A.2}$$

where y is a vector of time series with the magnitude of the transient pressure, λ is a factor that depends on the distance from the node to the source, and d is a vector for the corresponding damping factor. Once both factors are calculated, the values of the transient pressure simulation are updated as follows:

$$x_{new} = xk \tag{A.3}$$

$$y_{new} = yd + (1 - d)y_n \tag{A.4}$$

Figure A.1 presents a detailed example of the adjusted transient simulation results compared to the measured data and the original simulation at the sensor farthest from the transient source in Dataset 1. A comprehensive illustration of the results for all sensors in Dataset 1 is presented in Figure A.2, and a comparison of the original and adjusted transient model results used for threshold determination in both datasets is provided in Figure A.3.

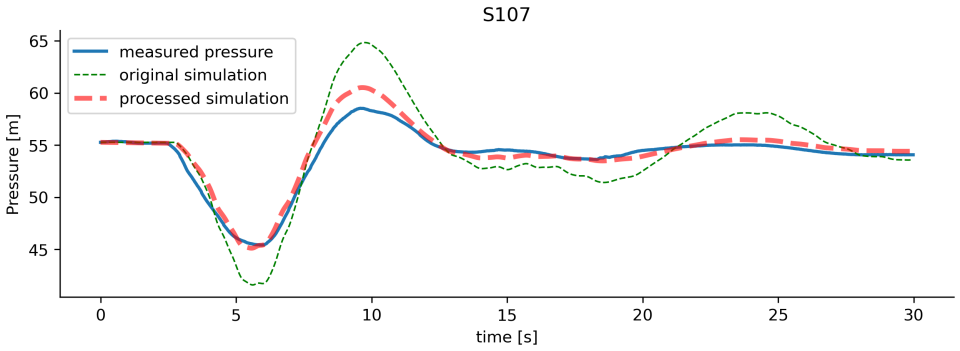


Figure A.1: Transient pressure simulation and post-processing of transient pressure Sensor 107

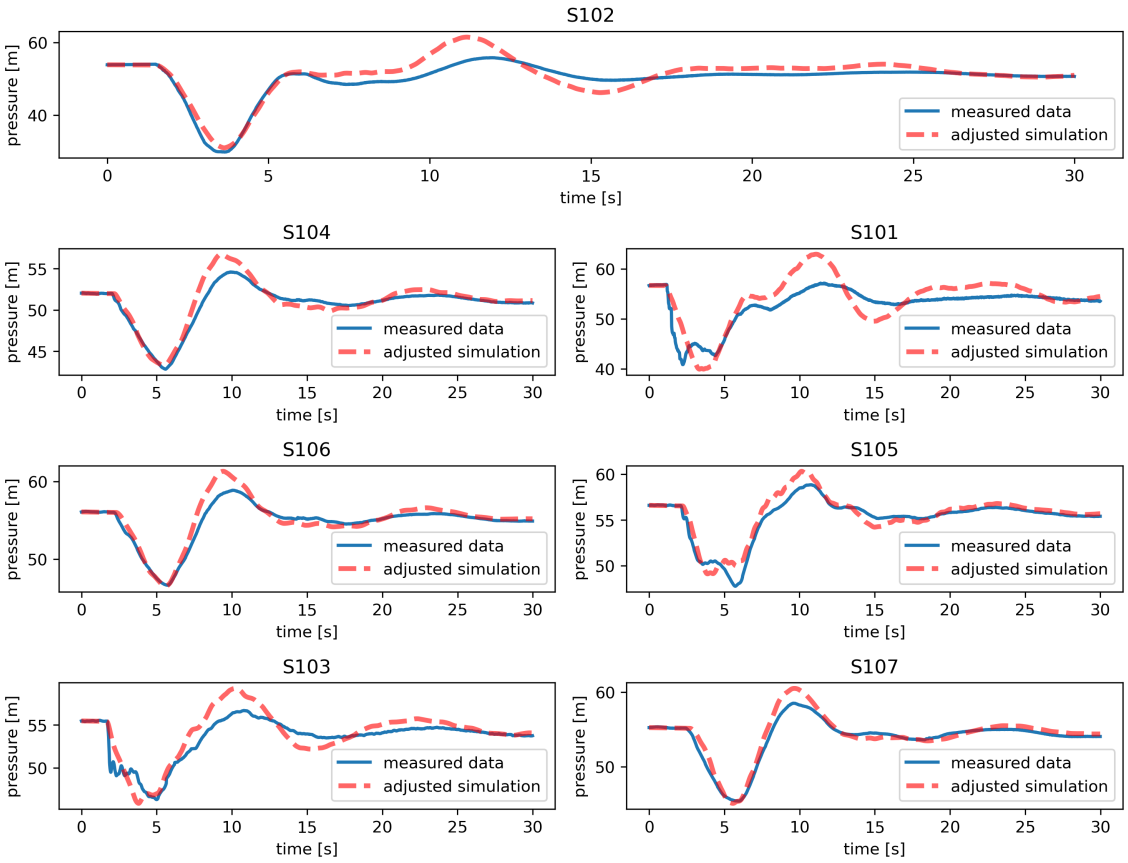
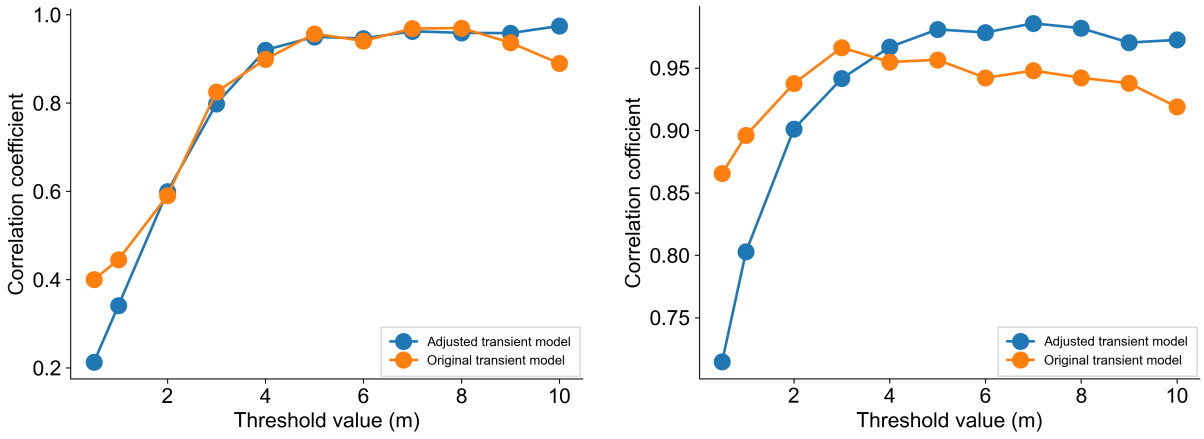


Figure A.2: Comparison adjusted transient model against actual measured data for all sensor readings in dataset 1



(a) (b)

Figure A.3: Comparison threshold estimation using original and adjusted transient model results. a) dataset 1. b) dataset 2

Appendix B

Additional testing scenarios

This appendix presents the results of the additional simulated testing scenarios 4 to 11, which are supplemental to the main three simulated scenarios presented in Chapter 8. Scenarios 10 and 11 (Figure B.7) are discussed in slightly more detail in the discussion section of Chapter 8. These scenarios (Scenarios 10 and 11) consider the case of two blocked pipes in the network. These scenarios use the same transient source conditions as scenario 2. The removed pipes in these scenarios correspond to pipes 122, and pipe 129. As in the previous scenarios, the transient source localisation algorithm is fed with the full network (considering the removed pipes).

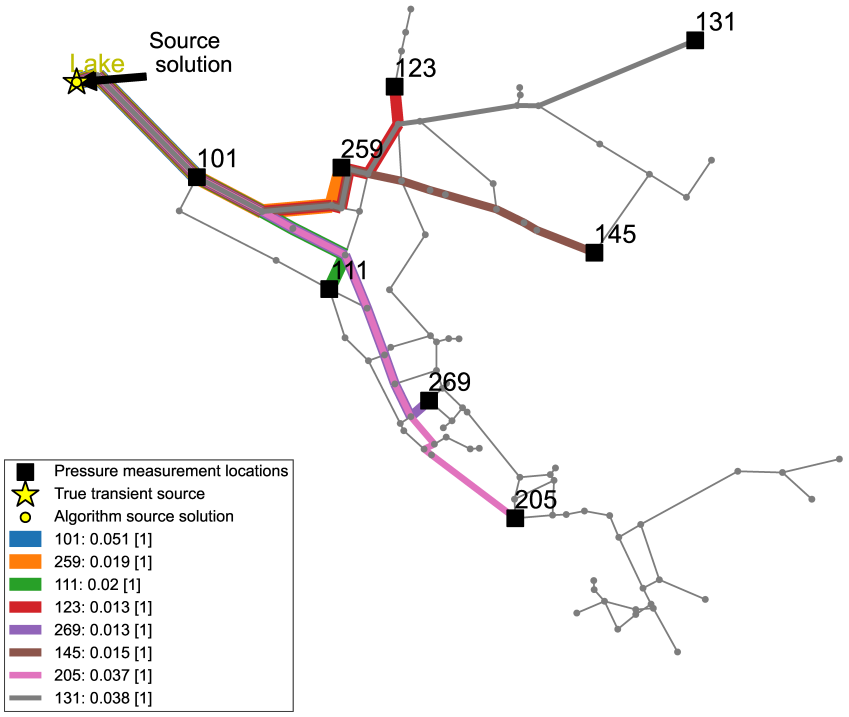


Figure B.1: Scenario 4. From: Jara-Arriagada et al. (2024).

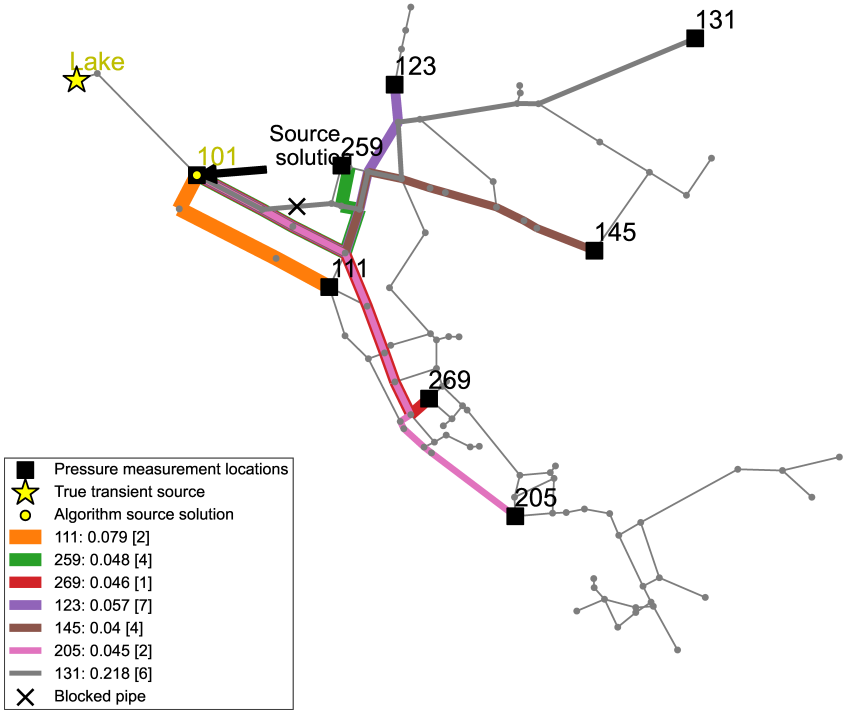


Figure B.2: Scenario 5. From: Jara-Arriagada et al. (2024).

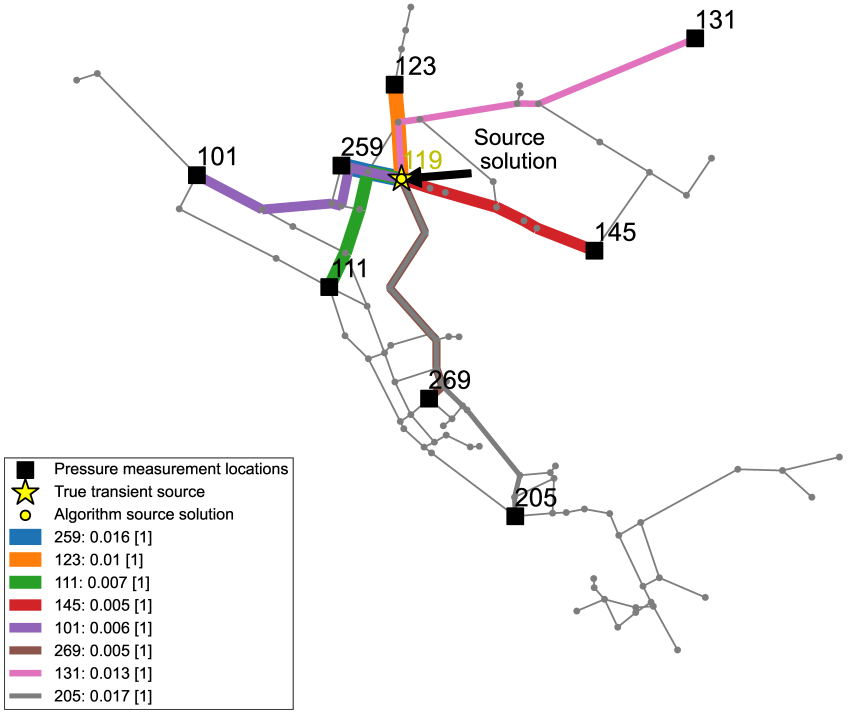


Figure B.3: Scenario 6. From: Jara-Arriagada et al. (2024).

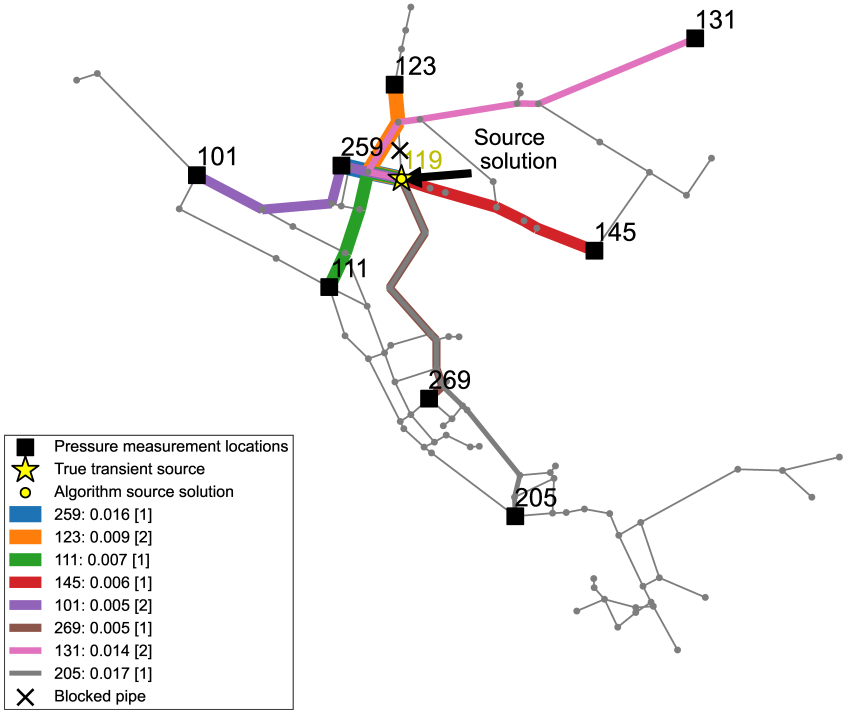


Figure B.4: Scenario 7. From: Jara-Arriagada et al. (2024).

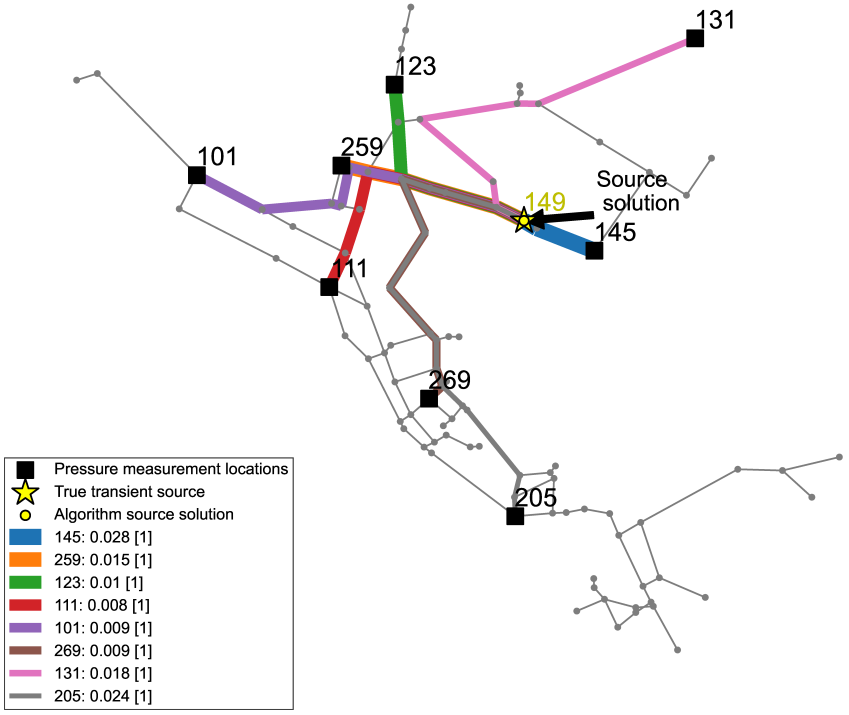


Figure B.5: Scenario 8. From: Jara-Arriagada et al. (2024).

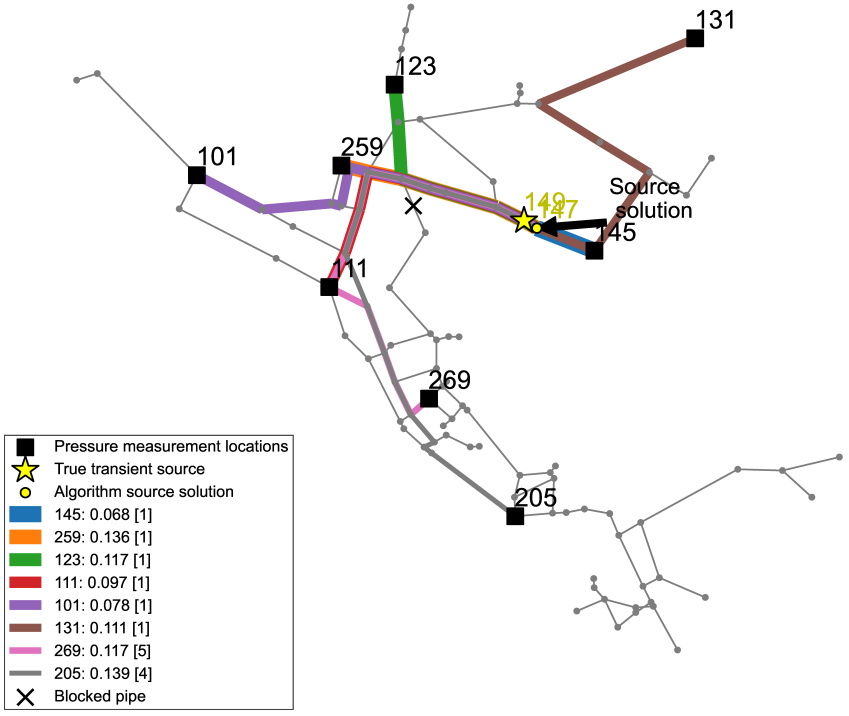
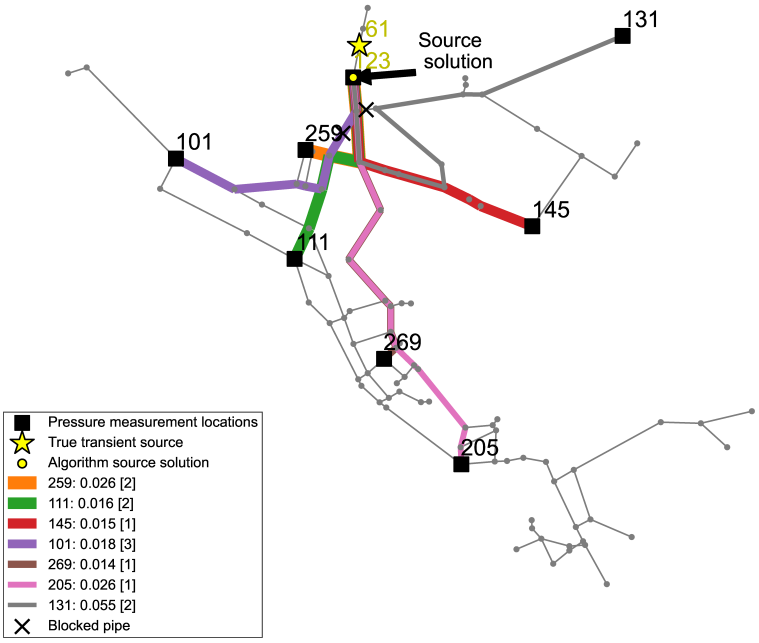
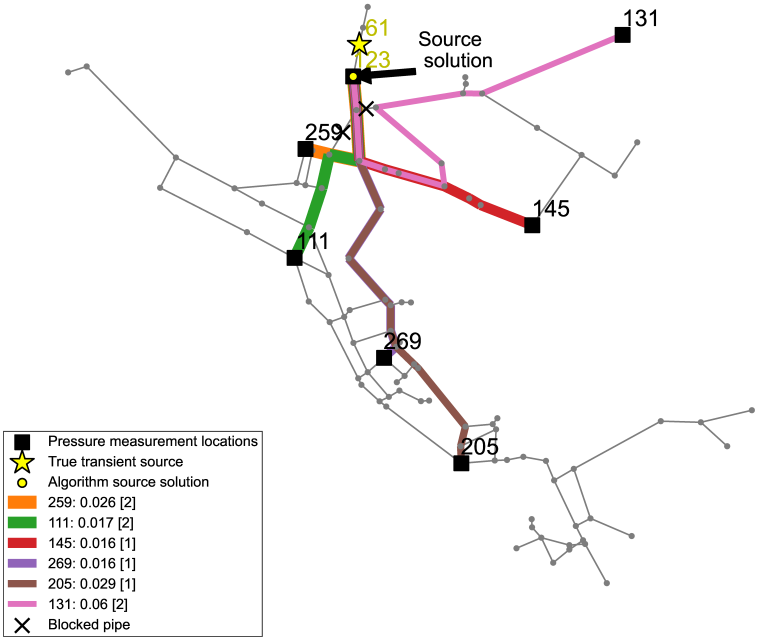


Figure B.6: Scenario 9. From: Jara-Arriagada et al. (2024).



(a)



(b)

Figure B.7: Additional testing scenarios. a) Scenario 10. Solution considering all sensors b) Scenario 11. Solution after removal of sensor 101. From: Jara-Arriagada et al. (2024).

Appendix C

Details of the weight function

This appendix provides a concise explanation of the weight function defined in Equation (8.6) in Chapter 8 to account for uncertainties in wave propagation and arrival times. As presented in section 8.2.1, the purpose of the optimisation approach is to minimise the error between measured and analytical times of arrivals of pressure waves. For this formulation, it is assumed that the wave speed is constant across a path and that the times of wave front arrivals can be precisely determined. However, because of potential attenuation of the pressure front head with distance, it can be difficult to precisely determine the exact point in time when the wave front arrives at a specific sensor location. As a result, measured times of wave arrivals are, in most cases, longer than analytical times (see also Table D.1 in Appendix D). This can have implications on the optimisation approach, which, in order to minimise errors, might select paths of higher orders (i.e. corresponding to longer propagation times) than the actual paths. One way to address this problem is to add a weight function to adjust analytical times of wave front arrivals.

In this case, the weight function presented in Equation (8.6) increases the time of wave propagation in a path by a factor of α, which is a way to deal with potential uncertainties in sensor measurements. The factor α is also multiplied by $\sqrt{l_p}$, where l_p indicates the order of the path p. Therefore, the higher the order of a path, the larger the value of the term $\alpha\sqrt{l_p}$. The effect of the weight function can be visualised with the example provided

in Figure C.1 and Figure C.2. Figure C.1 illustrates a minimal example network in which a transient is generated at node 0 and measured at nodes 1 and 2. This network also has a blockage on the pipe marked with a cross, which is not reflected by the hydraulic model. The thicker green lines show the actual propagation paths from the transient source to the sensors. Figure C.2 shows the potential measured times that can be recorded by the sensors and the analytical times that can be obtained together with the additional time provided by the weight function.

On the left hand side of Figure C.2, the potential measured times that the transient wave takes to reach sensors 1 and 2, denoted by t_1 and t_2, are represented in blue. The sensors are time-synchronised and the time at which they detect a transient wave depends on the start time at which the sensors were set. For example, the blue bars in Figure C.2 represent the magnitude of potential measured times of the pressure wave propagation to the sensors. These measured times will not be the same as the actual time that the pressure wave took to propagate from the source to the sensors, unless the sensors started to record time at the exact time the transient started. The offset between measured and actual times of arrival does not pose a problem, since the optimisation algorithm focuses on minimising differences of differences between times of arrival Δ against differences of analytical times of arrival δ. In other words, the optimisation algorithm will aim at selecting the analytical k-fastest paths (times) that, in combination, generate a close enough alignment between δ and Δ (bottom part of Figure C.2). As the reader can observe on the right hand side of Figure C.2, the weight function increases the analytical times, so that selecting higher order paths to compensate for errors in measured pressure wave arrival times becomes less necessary.

A main drawback of this approach is the assumption that the error in the measured time of wave front arrival always increases with the distance from the source to the sensor. However, if the determination of wave arrival times are precisely made for all sensors (including sensors farther away from the source), the weight could affect the performance of the localisation algorithm. Therefore, improvements for the weight function should be investigated in further research.

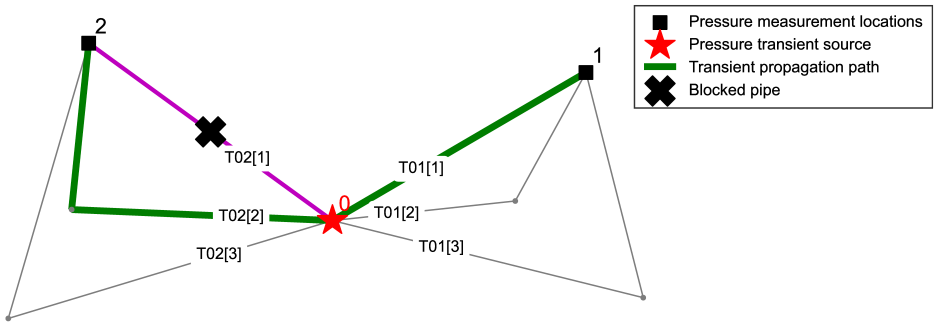


Figure C.1: Example of the different paths to be analysed by the algorithm. Path with a cross is blocked in real network, but available for consideration by the algorithm. From: Jara-Arriagada et al. (2024).

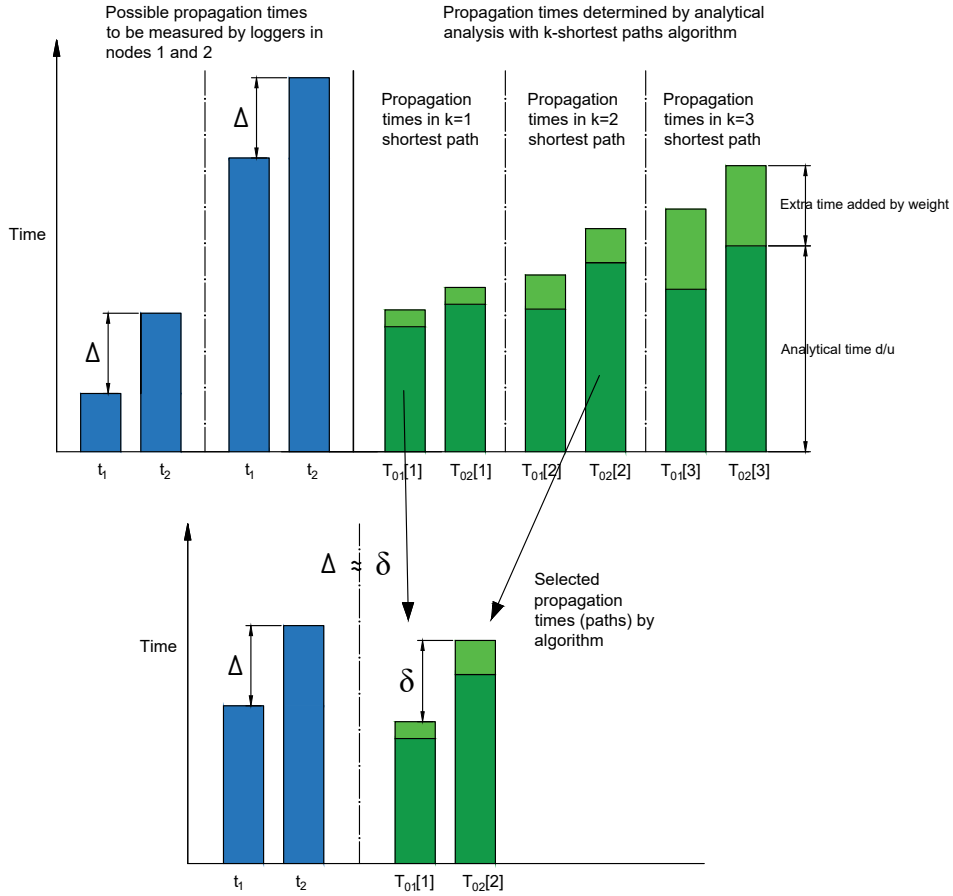


Figure C.2: The purpose of the weight function on additional time added to analytical times. From: Jara-Arriagada et al. (2024).

Appendix D

Measured (from simulation) vs analytical pressure wave arrival times

This appendix presents in Table. D.1 a quantitative example of the comparison between the time of wave propagation measured at the sensor locations after being simulated in TSNet and analytical pressure wave arrival times. The measure time of pressure wave propagation were determined by setting a threshold of -0.0045 in the change of gradient of the pressure time series. The analytical time is determined by dividing the length of the path by the pressure wave speed. The table also presents the additional time added by the weight function and how the difference between "measured" times and analytical times improve (see Table. D.1).

Table D.1: Comparison measured (from transient simulation) vs analytical pressure wave arrival times for scenario 1. From: Jara-Arriagada et al. (2024).

Sensor	Measured wave propagation time	Analytical wave propagation time	time difference	additional time by weight function	time difference with additional time
269	1.29617	1.23596	0.06021	0.07416	-0.01395
259	1.80943	1.70561	0.10382	0.10234	0.00148
111	2.14710	2.0320	0.11510	0.12192	-0.00682
123	2.29568	2.15646	0.13922	0.12939	0.00983
205	3.03856	2.88544	0.15312	0.17313	-0.02001
145	3.37623	3.18853	0.18769	0.19131	-0.00362
101	3.61935	3.39852	0.22083	0.20391	0.01692
131	5.01056	4.70916	0.30140	0.28255	0.01885

* Weight function applied with a value of $\alpha = 0.06$

Appendix E

Permissions to reproduce own published work

Related publications (Jara-Arriagada & Stoianov, 2022),(Jara-Arriagada & Stoianov, 2024), and (Konstantinou et al., 2024) are published open access under a Creative Commons Attribution 4.0 International Licence. For the related publication (Jara-Arriagada & Stoianov, 2021), the publisher, Elsevier, grants authors the right to re-use their articles in their own new works, without permission from Elsevier, including by using their articles in a thesis or dissertation. For the related publication Jara-Arriagada et al. (2024), the Licence Agreement with the publisher, The American Society of Civil Engineers, is attached here.

CCC Marketplace

This is a License Agreement between Carlos Ingnacio Jara Arriagada ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	16-Jan-2025	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1567613-1		
ISSN	1943-5452	**Publisher**	American Society of Civil Engineers
		Portion	Page

LICENSED CONTENT

Publication Title	Journal of Water Resources Planning and Management	**Rightsholder**	American Society of Civil Engineers
		Publication Type	e-Journal
Article Title	Localization of Transient Pressure Sources in Water Supply Networks with Connectivity Uncertainty	**Issue**	4
		Volume	150
		URL	http://scitation.aip.org/wro/
Author / Editor	American Society of Civil Engineers		
Date	01/01/1983		
Language	English		
Country	United States of America		

REQUEST DETAILS

Portion Type	Page	**Rights Requested**	Main product and any product related to main product
Page Range(s)	1-15		
Total Number of Pages	15	**Distribution**	Worldwide
Format (select all that apply)	Print, Electronic	**Translation**	Original language of publication
Who Will Republish the Content?	Academic institution	**Copies for the Disabled?**	No
Duration of Use	Life of current edition	**Minor Editing Privileges?**	Yes
Lifetime Unit Quantity	Up to 1,000,000	**Incidental Promotional Use?**	Yes
		Currency	GBP

NEW WORK DETAILS

Title	Methods for evaluating the impact of pressure transients on water distribution pipe breaks	**Institution Name**	Imperial College London
		Expected Presentation Date	2025-02-03
Instructor Name	Ivan Stoianov		

ADDITIONAL DETAILS

Order Reference Number	N/A

The Requesting Person / Organization to Appear on the License	Carlos Ingnacio Jara Arriagada

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Localization of Transient Pressure Sources in Water Supply Networks with Connectivity Uncertainty	**Title of the Article / Chapter the Portion Is From**	Localization of Transient Pressure Sources in Water Supply Networks with Connectivity Uncertainty
Editor of Portion(s)	Jara-Arriagada, Carlos; Ulusoy, Aly-Joy; Stoianov, Ivan	**Author of Portion(s)**	Jara-Arriagada, Carlos; Ulusoy, Aly-Joy; Stoianov, Ivan
Volume / Edition	150	**Issue, if Republishing an Article From a Serial**	4
Page or Page Range of Portion	1-15	**Publication Date of Portion**	2024-04-01

Marketplace Permissions General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach

not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such "clipping" or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit "publishing ventures" where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- a statement to the effect that such copy was made pursuant to permission,
- a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and
- a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also

ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service. Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery).*** The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or "clipping" does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the

prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Last updated October 2022