UNIVERSITY OF EXETER | Centre for Water Systems

COLLEGE OF ENGINEERING, MATHEMATICS AND PHYSICAL SCIENCES

Development of the Next Generation of Water Distribution Network Modelling Tools Using Inverse Methods

Submitted by

SOPHOCLES SOPHOCLEOUS

to the University of Exeter as a thesis for the degree of
Doctor of Engineering in Water Engineering
in January 2019

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……

I dedicate this thesis to my beloved grandmother, “Antroulla”

Abstract

The application of optimisation to Water Distribution Network (WDN) Modelling involves the use of computer-based techniques to many different problems, such as leakage detection and localisation. The success in the application of any model-based methodology for finding leaks highly depends on the availability of a well-calibrated model. Both leak detection and localisation, as well as model calibration are procedures that constitute the field of *inverse problems* in WDN modelling. The procedures are interlinked and dependent as when a leak is found and the model is updated its quality improves, while when a model is calibrated its ability to detect and localise leaks also improves. This is because both *inverse problems* are solved with the aim to mimic the behaviour of the real system as closely as possible using field measurements. In this research, both *inverse problems* are formulated as constrained optimisation problems.

Evolutionary Optimisation techniques, of which Genetic Algorithms are the best-known examples, are search methods that are increasingly applied in WDN modelling with the aim to improve the quality of a solution for a given problem. This, ultimately, aids practitioners in these facets of management and operation of WDNs. Evolutionary Optimisation employs processes that mimic the biological process of natural selection and “survival of the fittest” in an artificial framework. Based on this philosophy a population of individual solutions to the problem is manipulated and, over time, “evolves” towards optimal solutions. However, such algorithms are characterised by large numbers of function evaluations. This, coupled with the computational complexity associated with the hydraulic simulation of WDNs incurs significant computational burden, can limit the applicability and scalability of this technology across the Water Industry. In addition, the *inverse problem* is often “ill-posed”. In practice, the ill-posed condition is typically manifested by the non-uniqueness of the problem solution and it is usually a consequence of inadequate quantity and/or quality of field observations.

Accordingly, this thesis presents a methodology for applying Genetic Algorithms to solve leakage related *inverse problems* in WDN Modelling. A number of new procedures are presented for improving the performance of such algorithms when applied to the complex *inverse problems* of leak detection and localisation, as well as model calibration. A novel reformulation of the *inverse problem* is developed as part of a decision support framework that minimizes the impact of the inherent computational complexity and dimensionality of these problems. A search space reduction technique is proposed, i.e., a reduction in the number of possible solution combinations to the *inverse problem*, to improve its condition considering the accuracy of the available measurements. Eventually, this corresponds to a targeted starting point for initiating the search process and therefore more robust stochastic optimisations. The ultimate purpose is to increase the reliability of the WDN hydraulic model in localising leaks in real District Metered Areas, i.e., to reduce the number false positives. In addition, to speed up the leak search process (both computationally and physically) and, improve the overall model accuracy.

A calibrated model of the WDN is not always available for supporting work at distribution mains level. Consequently, two separate problem-specific methods are proposed to meet the abovementioned purpose: (a) a Leak Inspection Method used for the detection and localisation of leaks and; (b) a Calibration Method for producing an accurate average day model that is fit for the purpose of leak detection and localisation. Both methods integrate a three-step Search Space Reduction stage, which is implemented before solving the *inverse problem*. The aim is to minimize the number of decision variables and the range of possible values, while trying to preserve the optimum solution, i.e., reduce the *inverse problem* dimensionality. The search space reduction technique is established to generate a reduced set of highly sensitive decision variables. Eventually this is done to provide a viable, scalable technique for accelerating evolutionary optimisation applications in *inverse problems* being worthwhile on both academic and practical grounds.

The novel methodologies presented here for leak detection and localisation, as well as for model calibration are verified successfully on four case studies. The

case studies include two real WDN examples with artificially generated data, which investigate the limits of each method separately. The other two case studies implement both methods on real District Metered Areas in the United Kingdom, firstly to calibrate the hydraulic network model and, then, to detect and localise a single leak event that has actually happened. The research results suggest that leaks and unknown closed or open throttle valves that cause a hydraulic impact larger than the sensor data error can be detected and localised with the proposed framework which solves the *inverse problem* after search space reduction. Moreover, the quality of solutions can dramatically improve for given runtime of the algorithm, as 99.99% of infeasible solution combinations are removed, compared to the case where no search space reduction is performed. The outcomes of the real cases show that the presented search space reduction technique can reduce the search area for finding the leak to within 10% of the WDN (by length). The framework can also contribute to more timely detection and localisation of leakage hotspots, thus reducing economic and environmental impacts. The optimisation model for predicting leakage hotspots can be effective despite the recognized challenges of model calibration and the physical measurement limitations from the pressure and flow field tests.

Acknowledgements

I would like to take this opportunity to acknowledge all the people listed below for their contribution to this research, without whom this work would not have been accomplished.

Firstly, I would like to give very special thanks to my two EngD academic supervisors, Professor Dragan A. Savic and Professor Zoran Kapelan. I have learned a great deal from them over the last four years, and the work presented in this thesis would not have been possible without their help. These special thanks are conveyed not only for their continuous guidance and encouragement throughout this research, but also for their kindness, patience, perseverance and overall professionalism. With their solid and innovative thinking in doing research, they have provided me with invaluable advice.

My further thanks and appreciation go to my two EngD industrial supervisors, Mr Paul Sage and Mr Chris Gilbert. They provided me valuable technical support and advice from the practitioner's side in the field of Water Distribution Network Modelling. In addition, for their kindness and encouragement throughout my research, as well as for the time they spent listening to my research presentations and reading my long reports.

I would like to acknowledge all the support I got from the Clean Water Hydraulic Modelling Team of Severn Trent Water and especially Yibo Shen and Dave Andrews for their kind assistance in gathering resources for my research, the hydraulic models and field data.

In addition, I would like to thank to Professor Orazio Giustolisi, Assistant Professor Luigi Berardi and Associate Professor Danielle Laucelli from the Technical University of Bari. They have been my mentors at the early stages of my research career and from them I learnt many invaluable skills.

I would like to thank all the incredibly dedicated and talented members of the STREAM Industrial Doctorate Centre and the Centre for Water Systems, with whom I had the pleasure of working, for creating an inspiring research environment, and the friends that I have met over the last four years.

Finally, I wish to express my genuine gratitude to my parents, “Christina” and “Spyros”, for their enduring support over the period that I have been working towards this thesis. They have always been a great motivation for me in every step of my life. In addition, I would not be where I am today without the loving support of my grandfather, “Sophocles”. His encouragement, friendship and love made the journey towards this goal much easier. My gratitude extends to, my two beloved siblings, “Constantinos” and “Eleonora”, for the never-ending love, constant support and encouragement.

Sophocles Sophocleous
Exeter
December 2018

Table of Contents

Abstract 3
Acknowledgements 6
Table of Contents 8
List of Figures 13
List of Tables 16
List of Abbreviations 17
List of Notations 20
CHAPTER 1 Introduction 27
1.1 Background 27
1.2 Motivation for Research 29
1.3 Aims and Objectives of Research 31
1.4 Thesis Layout 33
1.5 Related Publications 35
CHAPTER 2 Inverse Problems in Water Distribution Network Modelling 38
2.1 Introduction 38
2.2 The Inverse Problem 39
2.2.1 Background 39
2.2.2 Problem Formulation 40
2.2.3 Well-Posed Vs Ill-Posed Inverse Problem 42
2.2.4 Practical Considerations 45
2.2.5 Strategy for solving Inverse Problems 46
2.3 Optimisation and Inverse Problems 48
2.3.1 Optimisation Problem Formulation 48
2.3.2 Search Space of Optimisation Problem 49
2.3.3 Genetic Algorithm Optimisation 51
2.3.4 Optimisation Methods Comparison 53
2.3.5 Search Space Exploration Enhancement 55
2.4 Water Distribution Network Model Calibration 57
2.4.1 Introduction 57
2.4.2 The Calibration Procedure 59

2.4.3 Data Collection Process 59
2.4.4 Calibration Approaches 60
2.4.5 Model Performance Criteria and Validation 63
2.5 Leak Detection and Localisation in Water Distribution Networks 65
2.5.1 Introduction 65
2.5.2 The Leak Detection and Localisation Procedure 67
2.5.3 Externally-Based Methods 68
2.5.4 Internally-Based Methods 71
2.5.4.1 Non-Mathematical Modelling 71
2.5.4.2 Mathematical Modelling 72
2.5.4.2.1 Leakage Modelling 72
2.5.4.2.2 Inverse Analysis Methods 73
2.5.4.2.3 Residual Analysis Methods 79
2.6 Concluding Remarks 81
CHAPTER 3 Search Space Reduction of Inverse Problems in Water Distribution Network Modelling 85
3.1 Introduction 85
3.2 Decision-Support Framework Overview 86
3.3 Inverse Problem Simplification 88
3.3.1 Introduction 88
3.3.2 Leakage Nodes 88
3.3.3 Valve Components 89
3.3.4 Pipe Roughness Groups 91
3.4 Parameter Sensitivity Analysis 92
3.4.1 Introduction 92
3.4.2 Minimum Detectable Nodal Leakage 93
3.4.3 Detectable Valve Locations 94
3.4.4 Detectable Pipe Roughness Coefficients 95
3.5 Search Space Optimisation 97
3.5.1 Introduction 97
3.5.2 Leak Locations and Range of Flow Values 97
3.5.3 Closed Valves 103
3.5.4 Pipe Roughness Calibration Groups 105
3.6 Inverse Problem Solving 107

3.7 Summary and Conclusions 108
CHAPTER 4 Methods for Leakage Detection and Localisation, and Model Calibration based on Search Space Reduction 111
4.1 Introduction 111
4.2 The Overall Leakage Detection and Localisation Process 113
4.2.1 Introduction 113
4.2.2 Sensor Data 116
4.2.3 Internally-Based Method 117
4.2.4 Externally-Based Method 117
4.3 Sensor Data Pre-Processing 118
4.3.1 Introduction 118
4.3.2 Sensor Data Quality 119
4.3.3 Data Pre-Processing 120
4.4 Leakage Inspection Method 123
4.4.1 Overview 123
4.4.2 Assumptions 125
4.4.3 Search Space Reduction Stage 125
4.4.4 Leakage Detection and Localisation Stage 126
4.5 Calibration Method 126
4.5.1 Overview 126
4.5.2 Data Pre-Processing Stage 129
4.5.3 Profile Calibration Stage 135
4.5.4 Search Space Reduction Stage 139
4.5.5 Component Calibration Stage 139
4.6 Summary and Conclusions 141
CHAPTER 5 Application of the Leakage Inspection and Calibration Methods to Case Studies 144
5.1 Introduction 144
5.2 Semi-Real Applications 146
5.2.1 Case Study SR1 146
5.2.1.1 Experimental Design 146
5.2.1.2 WDN Hydraulic Model Description 147
5.2.1.3 Leak Localisation Problem Set Up 148
5.2.1.4 Overall Inverse Problem Reduction 149

5.2.1.5 Leak Detectability 151
5.2.2 Case Study SR2 153
5.2.2.1 Experimental Design 153
5.2.2.2 WDN Hydraulic Model Description 156
5.2.2.3 Calibration and Leak Localisation Problem Set Up 157
5.2.2.4 Calibration Performance 160
5.2.2.5 Leak Localisation Accuracy 163
5.3 Real District Water System Applications 168
5.3.1 Case Study R1 168
5.3.1.1 System Overview 168
5.3.1.2 Available Data 169
5.3.1.3 Calibration Method Implementation 171
5.3.1.3.1 Data Pre-Processing 171
5.3.1.3.2 Profile Calibration 172
5.3.1.3.3 Component Calibration 174
5.3.1.4 Leak Inspection Method Implementation 177
5.3.1.4.1 Leak Search Area Reduction 177
5.3.1.4.2 Model Calibration 179
5.3.2 Case Study R2 181
5.3.2.1 System Overview 181
5.3.2.2 Available Data 182
5.3.2.3 Calibration Method Implementation 184
5.3.2.3.1 Data Pre-Processing 184
5.3.2.3.2 Profile Calibration 184
5.3.2.3.3 Component Calibration 186
5.3.2.4 Leak Inspection Method Implementation 188
5.4 Evaluation of LIM and CM 191
5.4.1 Operational Benefits of LIM and CM 191
5.4.1.1 Computational Benefits of Search Space Reduction 193
5.4.1.2 Practicality of the LIM 198
5.4.1.3 Practicality of the CM 200
5.4.1.4 Model Calibration and Data Challenges 202
5.5 Summary and Conclusions 204

CHAPTER 6 Summary, Conclusions and Further Work Recommendations 207
6.1 Introduction 207
6.2 Summary of Present Work 208
6.2.1 Leak Detection and Localisation 209
6.2.2 Hydraulic Network Model Calibration 209
6.3 Research Contributions 210
6.3.1 Inverse Problem 210
6.3.2 Leak Detection and Localisation 211
6.3.3 Model Calibration 212
6.3.4 Water Distribution Network Modelling 212
6.4 Research Conclusions 213
6.5 Further Work Recommendations 216
6.5.1 Leak Detection and Localisation 216
6.5.2 Model Calibration 217
6.5.3 Optimisation Technique and Data Effect 218
APPENDIX A Leakage Inspection Tool 221
A.1 Tool Description 221
A.2 Prerequisites and Inputs - Outputs 222
A.3 End User and Process Time Requirements 223
APPENDIX B Calibration Tool 226
B.1 Tool Description 226
B.2 Prerequisites and Inputs - Outputs 227
B.3 End User and Process Time Requirements 227
Bibliography 231

List of Figures

Figure 2.1. Forward vs Inverse Problem. 40
Figure 2.2. The conditions of a well-posed inverse problem. 43
Figure 2.3. Example of Landscape in search space. 51
Figure 2.4. Search Space Reduction. 56
Figure 2.5. The Implicit Calibration Procedure (Savic et al. 2009). 62
Figure 3.1. Overview of the decision-support framework for solving inverse problems in WDN modelling based on search space reduction. 87
Figure 3.2. Reduction via Inverse Problem Simplification for the candidate leak locations. 89
Figure 3.3. Network representation of valves that lead to isolation. 90
Figure 3.4. Search Space Optimisation for Leakage Nodes. 102
Figure 3.5. Search Space Optimisation for Valve Components. 104
Figure 3.6. Search Space Optimisation for Pipe Groups. 107
Figure 4.1. The leak localisation and model calibration feedback loop. 112
Figure 4.2. The proposed novel procedure used for leakage detection and localization. 115
Figure 4.3. Schematic Representation of Data Pre-Processing. 122
Figure 4.4. Overview of the Leak Localisation methodology framework. 124
Figure 4.5. Overview of the Calibration Methodology framework. 129
Figure 4.6. Example of the data pre-processing method applied on the available flow measurements. 132
Figure 5.1. The SR1 Case District Water System. 148
Figure 5.2. Percentage of the total WDN nodes representing potential leak locations that are left for the use with the detection and localisation methodology based on different leak sizes (figure legend) and noise levels. . 150
Figure 5.3. Mean Percentage Pressure Difference for each leak flow scenario relative to leak-free case. 151
Figure 5.4. The general process flow chart for the experimental design. 154
Figure 5.5. The SR2 Case District Water System. 157
Figure 5.6. The generated inlet flow raw datasets for each DF scenario. 159

Figure 5.7. The calibration dataset for scenario DF5. 160
Figure 5.8. Average distance between all the detected leaks compared to the actual leak locations, for each calibration and demand fluctuation scenario. . 164
Figure 5.9. Distance between each detected leak location relative to the actual leak locations, for each calibration and demand fluctuation scenario. 165
Figure 5.10. Comparison of reported leak size for each CAL scenario, averaged for scenarios DF2 – DF5. 166
Figure 5.11. Comparison of the normalized objective function value (0-worst, 1-best) achieved after leak localisation for each CAL scenario as a result of changing demand fluctuation. 168
Figure 5.12. The R1 Case District Water System. 169
Figure 5.13. The time series field record of the inlet flow meter and pressure data from the closest sensors to the leak location. 170
Figure 5.14. Comparison of the resulting average day dataset and the calibration dataset after the DPP stage, relative to the raw observations. 172
Figure 5.15. Comparison of the updated hydraulic model before and after Demand Profile Calibration as well as relative to the calibration dataset. 172
Figure 5.16. Comparison of the observed and simulated heads for the uncalibrated model and optimised calibration solution during MNF at hour 3 (top) and the morning peak at hour 8 (bottom). 173
Figure 5.17. Part II optimisation analysis outcome for the different leak scenarios in the WDN (0-worst, 1-best). 177
Figure 5.18. Leakage Hotspot Map after the optimisation analysis. 178
Figure 5.19. The Flow Differences before and after leakage detection. 179
Figure 5.20. Comparison of the observed and simulated heads for the optimised leakage detection solution during MNF at hour 4. 180
Figure 5.21. The R2 Case District Water System. 181
Figure 5.22. The Field Record with the measurements used for model calibration and for finding the leak. 183
Figure 5.23. Comparison of the resulting average day dataset and the calibration dataset after the DPP stage, relative to the raw observations. 184
Figure 5.24. Comparison of the updated hydraulic model before and after Demand Profile Calibration as well as relative to the observations. 185

Figure 5.25. The detected correct status of the identified valves following the model calibration. 186
Figure 5.26. Comparison of the observed and simulated heads for the uncalibrated model and optimised calibration solution during MNF at hour 3 (top) and the morning peak at hour 8 (bottom). 187
Figure 5.27. Leakage Hotspot Map after the optimisation analysis. 189
Figure 5.28. The Flow Differences before and after leakage detection. 190
Figure 5.29. Comparison of the observed and simulated heads for the optimised leakage detection solution during MNF at hour 4. 191

List of Tables

Table 2.1. Summary of Implicit Calibration Procedures. 64
Table 2.2. Calibration criteria for flow and pressure. 65
Table 2.3. Main characteristics of externally-based techniques. 70
Table 2.4. Summary of Leak Detection and Localisation Procedures based on Inverse Analysis. 78
Table 2.5. Summary of Leak Detection and Localisation Procedures based on Forward Analysis. 81
Table 2.6. Main characteristics of internally-based techniques. 82
Table 5.1. Overview of the presented Case Study Applications..... 145
Table 5.2. Comparison of geographical distances (m) of the detected leak from the assumed (true) leak location for each flow scenario and noise level. 152
Table 5.3. Overview of the tested desktop calibration experiments. 158
Table 5.4. The Candidate Grouping after Step 1 of SSR Stage. 162
Table 5.5. The Candidate Grouping after Step 1 of SSR Stage. 175
Table 5.6. Simulated Pressures relative to the calibration accuracy criteria. . 176

List of Abbreviations

AC	Asbestos Cement
ACO	Ant Colony Optimisation
CAL	Calibration
CLONALG	Clonal Selection
CC	Component Calibration
CI	Cast Iron
CM	Calibration Method
CSV	Comma Separated Values
DDA	Demand-driven Analysis
DDS	Dynamically Dimensioned Search
DE	Differential Evolution
DF	Demand Fluctuation
DMA	District Metered Area
DPRC	Detectable Pipe Roughness Coefficients
DPC	Demand Profile Calibration
DPP	Data Pre-Processing
DVL	Detectable Valve Locations
EA	Evolutionary Algorithm
EPS	Extended Period Simulation
FC	Friction Coefficient
fmGA	Fast messy Genetic Algorithm
GA	Genetic Algorithm
GN	Gauss-Newton
GPRS	General Packet Radio Service
GRG	General Reduced Gradient
GSM	Global System for Mobile communications
HMO	Honeybee Mating Optimisation
HPPE	High Performance Polyethylene
HS	Harmony Search
IPS	Inverse Problem Solving
IPSI	Inverse Problem Simplification

KF	Kalman Filter
LDL	Leak Detection and Localisation
LF	Leak Flow
LIM	Leak Inspection Method
LLC	Lumped Leak Coefficient
LLI	Leak Location Index
LM	Levenberg-Marquardt
LOC	Leak Orifice Coefficient
LOWESS	Locally Weighted Polynomial Regression
LP	Leak Parameters/Linear Programming
LS	Link Status
LSS	Leak Signature Space
MDNL	Minimum Detectable Nodal Leakage
ME	Maximum Error
MF	Model Falsification
MNF	Minimum Night Flow
ND	Nodal Demand
NDM	Nodal Demand Multiplier
ODBC	Open Database Connectivity
ODS	One Dimensional Search
OF	Objective Function
PC	Profile Calibration
PD	Pipe Diameter
PDD	Pressure Driven Demand
PLI	Pipe Location Index
PPC	Pressure Profile Calibration
PRV	Pressure Reducing Valve
PS	Pump Speed
PSA	Parameter Sensitivity Analysis
PSO	Particle Swarm Optimisation
QN	Quasi Newton
R	Real
RC	Generalised Resistance Coefficient
RM	Residuals Matrix

RMSE	Root Mean Squared Error
SA	Simulated Annealing
SAE	Sum of Absolute Error
SCEM	Shuffled Complex Evolution Metropolis
SMS	Short Message Service
SR	Semi-Real
SSE	Sum of Squared Error
SSO	Search Space Optimisation
SSR	Search Space Reduction
TDEM	Total Demand
VR	Valve Resistance
VS	Valve Setting
WDN	Water Distribution Network
WSAE	Weighted Sum of Absolute Errors
WSSE	Weighted Sum of Squared Errors
UK	United Kingdom

List of Notations

a	Pressure exponent of leak orifice equation
$Accuracy_p$	Sensor device reading percentage error
$Accuracy_Q$	Flow device reading percentage error
$B_{j,p}$	Baseline demand for demand type p at junction j
BS_k	Baseline head, setting or speed at model component k
c_i	Emitter coefficient depicting the leak orifice area at node i
c_i^n	Emitter coefficient for node i for the possible leak n
CM_{DPRC}	Number of hydraulic simulations required to determine all the detectable pipe roughness coefficients
CM_{DVC}	Number of hydraulic simulations required to determine all the detectable valve locations
CM_{MDNL}	Number of hydraulic simulations required for the MDNL Analysis
CM_{PartI}	Number of hydraulic simulations required to detect the total water losses in Part I of Step 3 in SSR
CV_k^v	Index of valve k corresponding to possible closed valve v
d	Day
d_i	Time distance from $y_{o,k}$
d_z	User-specified range of possible emitter coefficient values for optimisation analysis z
$DPRC_{pg,pm}$	Set of possible roughness values for detectable pipe group pg of material pm
DV	Number of decision variables for the optimisation problem
F	Number of flow metered pipes
f_i	Roughness coefficient for pipe or pipe group i
fi_z	Fitness improvement after optimisation analysis z
FI^g	Set of fitness improvement values for pipe group scenario g
g	Pipe group scenario
G	User defined generations number parameter of the Genetic Algorithm
GS	Ranked list of sensitive pipe groups

$Ho_s(t)$	Observed head at node s at time t
$Hpnt$	Pressure device reading error
$Hs_s(t)$	Simulated head at node s at time t
J	Set of potential leak locations for any possible leak
k	Candidate valve location
K	Set of possible values for any possible leak
l	Candidate leak location
LN_i^n	Index for node i for the possible leak n
m	Sensor device location
$m_{j,t}$	Demand multiplier for junction or junction group j at time t
M	Number of sensor devices
$MDNL$	Set of Minimum Detectable Nodal Leakage flow values for each candidate location l
n	Possible leak
NEC	Number of emitter coefficients that result after the user specifies the maximum possible flow before the MDNL analysis
NCN	Number of candidate nodes after Step 1 of SSR
$NCVC$	Number of candidate valve components after Step 1 of SSR
$NGroups$	Number of pipe group scenarios
NI	Total number of candidate pipes or pipe groups
$NIndex$	Number of candidate nodes for any possible leak n
NJ	Total number of candidate junctions or junction groups
NK	Total number of candidate valves
$NLdup^n$	Number of duplicate nodes identified as leakage emitters
$NLdup^v$	Number of duplicate valves identified as closed
$NLeak$	Total number of leaks to be identified
$Npattern$	Number of patterns to be calibrated
NPG	Number of candidate pipe groups after Step 1 of SSR
NRC	Number of candidate roughness coefficient values that correspond to the candidate pipe groups
$NStep$	Number of time steps for each pattern p
$NValve$	Number of closed valve scenarios

OF_0	Objective function value for the boundary conditions
OF_z	Objective function value for optimisation analysis z
P	Pressure at any node
$\widehat{p_{0s}}$	Pressure at node s for a leak-free scenario
p_i	Pressure at node i
$p_{k,s}$	Pressure response at node s for a change in the status of valve k
$p_{Q,l,s}$	Pressure response at node s for leak flow rate Q at candidate location l
$p_{\lambda_{pg,pm},s}$	Pressure response at node s for a change in roughness value λ of pipe group pg of material pm
pa_{pm}	Pipe age sub-class for material pm
PA_{pm}	Number of pipe age sub-classes for material pm
PA	Number of pipe age groups
pg	Pipe group
$pg_{pa_{pm}}$	Pipe diameter group for age sub-class pa for material pm
PG	Number of pipe diameter groups
$PG_{pa_{pm}}$	Number of pipe diameter groups for age sub-class pa for material pm
$P_k(t)$	Pressure at the model component k at time t
pm	Pipe Material Class
PM	Number of pipe material classes
PM_t^p	Pattern multiplier for pattern p at time step t
$p_{Q,l,s}$	Pressure at node s for leak flow Q at candidate location l
$\overline{P_s}$	Average observed pressure at sensor nodes
P^p	Set of pattern multipliers for pattern p
$PPat_k(t)$	Pattern multiplier for the baseline head, setting or speed BS_k of model component k at time t
PS	User defined population size for the number of solutions generated by the Genetic Algorithm
Q	Leakage flow rate
Q_i	Leakage flow rate at node i

$Q_j(t)$	Total demand at junction j at time t
$\bar{Q}_{in}$	Average global system demand
q_l	Leakage flow at candidate leak location l
$Q_{\overline{c^n}}$	Detected leak flow
$Qmin_l$	Minimum Detectable Nodal Leakage flow value for candidate location, l
$Qo_f(t)$	Observed flow in link f at time t
$QPat_{j,p}(t)$	Pattern multiplier for demand type p at junction j at time t
$Qs_f(t)$	Simulated flow in link f at time t
$Qpnt$	Flow device reading error
r	Residual error between measured and observed value
$r_k(t)$	Residual error for a change in the status of valve k
r_m	Residual error between measured and simulated value at location m
$r_{Q,l}(t)$	Residual error for leak flow rate Q at candidate location l
$r_{\lambda_{pg,pm}}(t)$	Residual error between the minimum nominal roughness value and the adjusted roughness coefficient $\lambda_{pg,pm}$ for pipe group pg of material pm
$RIndex$	Number of possible roughness values for pipe group g
RV_{pm}	Set of possible roughness values for pipe material class pm
$RV_{pg,pm}$	Set of possible roughness values for pipe group pg of material pm
s	Sensor node location
S	Number of sensor locations
$s_{k,t}$	status of valve k at time step t
$SS_{BaseCaseValves}$	Number of hydraulic simulations required to evaluate any possible solution to the base case optimisation problem for detecting the status of valves without any reduction in space
$SS_{BaseCaseLeaks}$	Number of hydraulic simulations required to evaluate any possible solution to the base case optimisation problem for leak detection and localisation without any reduction in space
$SS_{BaseCasePipes}$	Number of hydraulic simulations required to evaluate any

	possible solution to the base case optimisation problem for calibrating the roughness coefficient of pipe components without any reduction in space
$SS_{ClosedValves}$	Number of hydraulic simulations required to evaluate any possible solution to the optimisation problem in finding the maximum number of closed valves
SS_{LDL}	Number of hydraulic simulations required to evaluate any possible solution to the ultimate leak detection and localisation problem
SS_{PartII}	Number of hydraulic simulations required to evaluate any possible solution to the optimisation problem in Part II of SSR
SS_{PRCG}	Number of hydraulic simulations required to evaluate any possible solution to the optimisation problem in finding the maximum number of pipe roughness calibration groups
t	Time step
T	Total number of days considered in set of measurements
v	Closed valve scenario
V	Set of candidate valve locations
$VIndex$	Number of candidate valves for any possible closed valve
w_f	Weighting factor for observed flow in pipe f
$w_k(t)$	Tri-cube weight function at time step t
w_m	Weighting vector corresponding to set of measurements, y_o
w_s	Weighting factor for observed head at sensor s
x	Set of parameters for the inverse problem
x_i	Decision variable i of the optimisation problem
x_{SSR}	Reduced set of parameters for the inverse problem
$\vec{X}$	Set of decision variables for the optimisation problem
y_a	Actual value of measurement(s)
y_o	Sensor device reading(s)
y_o^*	Smoothed value for observation(s)
$y_{o,i}$	Nearest neighbour, i of $y_{o,k}$
$y_{o,k}$	Observation value to be smoothed
$\overline{y_{o,m}}(t)$	Average value of the readings from device m at time t

$y_{o,m,d}(t)$	Reading value from device m at time t on day d
$y_s(x)$	Model prediction(s) corresponding to measurement(s), y_o
z	Optimisation analysIs
Z	User defined number of optimisation analyses for the tested range of emitter coefficient values
ε	Sensor device error
λ_j^g	Roughness coefficient value for group of pipes g corresponding to index j
λ_{pm}	Roughness coefficient value for pipe material class pm
$\lambda_{pg,pm}$	Roughness coefficient value for pipe group pg of material pm
Λ	Number of possible roughness values for pipe group pg
λmin_{pm}	Minimum nominal roughness coefficient for pipe material pm

CHAPTER 1

Introduction

1.1 Background

Water is a valuable resource and essential for the human survival. It is used for drinking, food production, several social purposes and contributes to the economic growth of societies. However, less than 1% of the Earth's water is available for direct human consumption (USGS, 2011). This fraction is also unevenly distributed spatially and temporally (Gleick, 1995). Furthermore, demographic, socioeconomic and environmental factors, such as the accelerated population growth, unsustainable consumption patterns, pollution of underground and surface water, and extreme environmental fluctuations, put increasing pressure on the availability of fresh water resources. Thus, securing water supply for future generations is a compelling and critical issue. As a result, more emphasis is shifted towards managing demand by better utilising the water that is already available.

Water is supplied to consumers through water distribution networks (WDN). Considerable financial resources are allocated to the design and construction of a network that satisfies the supply and quality standards. On the other hand, a significant proportion of the water distributed is lost through leakage. Leakage in WDNs includes water that escapes from the pipe network by means other than through a controlled action (Ofwat, 2008a). According to a World Bank study, about 48 billion m^3 of water are lost annually from WDNs, costing water utilities approximately US$141 billion per year around the world (Kingdom et al., 2006; van den Berg, 2014).

In view of the concerning future water scarcity and environmental problems, water utilities have played a central role in conserving water through advances in leakage management. For example, leakage in England and Wales has been

reduced by 38% since the water industry privatization in 1989, from a peak of 5,155 to 3,183 Ml/day (Discover Water, 2018). Nevertheless, 35% of the saved water was achieved within the first 15 years and since then, the leakage reduction progress has been very slow (Environment Agency, 2008). Still, the total leakage reduction is equivalent to the daily needs of more than 14 million domestic customers (Ofwat, 2018). The overall progress in managing leakage was brought about by an investment in leak detection, repair and replacement of WDNs. Furthermore, as a result of managing pressures in WDNs and water mains rehabilitation programs. Yet, water infrastructure in various countries including the United Kingdom (UK) is ageing and new infrastructure is not being installed quickly enough to maintain current standards (Speight, 2015). Consequently, the lack of investment in pipe renewal will lead to an increase in pipe failure rates and greater levels of leakages in the future. Hence, water companies need to find solutions towards reducing risks and ensure more resilient systems.

Most water utilities have adopted operational strategies to address the leakage challenge. Nevertheless, they have limited effectiveness and do not address the root cause. Although plugging the leaks seems to be the most direct approach to solving the problem of water lost through leakage that is not a trivial task. As WDNs are buried, locating the leaks is extremely difficult. Many leaks do not manifest themselves in any visible way, such as water bubbling to the surface, and may go undetected indefinitely. Even when water does reach the surface, it may appear nowhere near the actual leak.

There are a number of technologies available for detecting leaks without digging, such as a variety of acoustic methods and step-wise testing of individual segments of the piping systems (Puust et al., 2010). However, these approaches are not effective in pinpointing leaks and still require fairly exhaustive examinations of the water systems. They are, therefore, expensive and time-consuming. Water suppliers generally seek to reach an “economic level of leakage”, i.e., ‘the level beyond which it is no longer economical to detect and repair leaks compared to other means of balancing supply and demand’ (Ofwat, 2008b). Nevertheless, even if water suppliers approach their “economic” level of leakage, they face further challenges. For example, the volume of water “wasted”

through leakage is becoming increasingly unacceptable in the eyes of the consumer. However, consumers are generally not willing to pay for an increase in charges to fund further reductions in leakage. As it is becoming increasingly difficult to find undiscovered leaks, the attention has turned to modelling technology that can detect and localise leaks.

Obviously, a mathematical model that can simulate WDN behaviour as closely as possible is of great importance to every WDN authority. Today, computer-based hydraulic simulators of the WDN are widely used by engineers for planning, design and operational purposes. These hydraulic models represent the real systems using mathematical equations to analyse the WDN behaviour and predict its response under a wide range of conditions (Walski et al., 2001). Therefore, it is important that the model represents the real state of the system to provide adequate solutions for various purposes (Walski, 2000). For example, a well-calibrated WDN hydraulic model can be used to detect and localise leaks.

1.2 Motivation for Research

Water distribution network hydraulic models are in widespread use by planners, water utility personnel, consultants and many others for the analysis, design, operation or maintenance of WDN. With the advent and subsequent evolution of this technology, engineers have been able to analyse the status and operations of WDN, as well as to investigate the impacts of proposed changes. Using a hydraulic model to identify areas of leakage has long been a goal for the practitioners. This challenge can be posed as an *inverse problem*. In an *inverse problem* in WDN modelling actual measurements are taken from the field and are compared with the outputs from simulated scenarios associated with the state of the WDN. The solution to the problem aims to determine the simulated scenario that yields the best match with the actual measurements and represents the real system situation, e.g., the size and the location of leaks. Solving the *inverse problem* for leak detection and localisation could use pressure and flow data obtained in the field to run an optimisation analysis using a pressure-driven hydraulic model to locate leaks for further investigation.

A considerable aspect of effective and meaningful network modelling is the calibration of models. This is a process in which model parameters regarding the state and status of the WDN are adjusted until the model matches (as closely as possible) the behaviour of the real WDN (Shamir & Howard, 1977). Calibration is an *inverse problem* related to WDN modelling, also called the "calibration problem". The process is normally employed to simulate pressures within an accuracy of ±1 metre relative to the observations. However, for supporting operational work at the distribution mains level, such as finding leaks, this is too coarse a criterion. Reasons for this involve undiscovered leaks, incorrect pipe state information and/or system anomalies that cause changes in the observed pressures which are smaller than the ±1m accuracy range of hydraulic models. For example, throttle valves that were accidentally left closed (or open) without being correctly included in the model raise significant modelling challenges. These are not updated in the Geographic Information Systems of the water utility and can be carried over in the water mains records used for model building and calibration. Except from that, traditionally, calibration was, and, in some cases still is treated as a manual task that is subject to the engineer's judgement. All the above mentioned factors cause a considerable effect on how accurately can the model simulate WDN hydraulics and, consequently, on the leak localisation process. On the other hand, relative to the traditional trial-and-error approaches, much better results can be achieved if the calibration problem is formulated as an optimisation problem (Savic & Walters, 1995).

Automated analysis and optimisation tools have been used to provide a mechanism for improving various facets of WDNs including, model calibration, and leakage detection and localisation (Puust et al., 2010). Such tools represent an attempt to provide assistance to practitioners through the means of intelligent, knowledge-based techniques. An optimisation model helps engineers find a solution for a given problem. When linked with a pressure-driven hydraulic model it can be used to generate multiple solutions that are, then, simulated and their corresponding impact on the WDN is recorded. Through a simulation-optimisation approach the quality of a solution generated by the optimisation model can be evaluated. On the other hand, the application of optimisation to WDNs is generally characterised by extended runtimes owing to the

computational load imposed by the simulation model. Determining the WDN hydraulics numerically is iterative in nature and the time taken to run a hydraulic simulation is largely dependent on the size and configuration of the network model itself. In addition, a significant problem associated with the quality of the generated solutions and shortfall in the efficacy of real-life WDN hydraulic models is the lack of measurements in the WDN. This can lead to the formulation of an ill-posed *inverse problem*, characterised by non-uniqueness and instability of solutions. So far, not much research effort has been put into tackling this problem.

Evolutionary Algorithms are search methods that are often used for optimising complex problems (Hassan et al., 2005). These methods have been successfully applied in WDN modelling in the past with the aim to improve the quality of a solution for a given problem (Mala-Jetmarova et al., 2015). However, they require a large number of solution evaluations and appear to be ill-suited for practical WDN applications. Their popularity in research rose due to their ability to converge rapidly on an optimal or near-optimal solution, whilst having analysed a mere fraction of the total solution search space. Despite inexorable improvements in computer power, there is still a need to improve the performance of these algorithms, to permit practical WDN applications. Such search methods need to be utilized with acceptable computational efficiency and effectiveness to meet the demands of the more complex optimisation problems faced in real life.

1.3 Aims and Objectives of Research

A family of population-based search algorithms known as evolutionary algorithms has been extensively considered in the field of WDN modelling (Goldberg, 1989). However, very few of the algorithms have received widespread acceptance in the commercial applications. This is because most algorithms require high number of function evaluations and computational time to solve even a simple problem. The aim of this research work is to develop novel approaches for simplifying the deployment of optimisation techniques in WDN modelling applications, such as model calibration and leak detection and localisation. The developed methodology aims to detect the state of assets and their status changes in a fast and reliable way. The ultimate purpose is to increase the reliability of the WDN

hydraulic model in localising leaks in real District Metered Areas (DMAs), speed up the search for leaks (both computationally and physically) and improve the model accuracy.

Novel approaches are proposed in the thesis to improve the condition of the *inverse problem*. They are aimed at improving the efficiency of the evolutionary optimisation algorithms by reducing the search space, i.e., the number of possible solutions to the *inverse problem*. This ensures that the algorithms minimize computational wastage by avoiding the evaluation of solutions do not cause impact on the model fitness. It also promotes the efficient convergence of the population to an optimal solution. Considering the high computational burden of optimisation techniques, accelerating their performance is of great importance.

The main objective of this thesis is to develop a decision-support framework for solving the *inverse problem* in WDN modelling, including calibration and leak detection and localisation, being worthwhile on both academic and practical grounds. The *inverse problem* formulation is solved using an optimisation technique linked to a pressure-driven hydraulic model. It is hypothesised that the search reduction approach will enhance the efficiency of the optimisation algorithm for solving *inverse problems* in WDN modelling, such as calibration and leak detection and localisation.

More specifically, the detailed objectives are as follows:

(1) To develop a universal approach for reducing the search area for a detecting leak/burst event or unknown throttle valve status that can be applied to any DMA network. The approach will take into account the sensor configuration and sensitivity of pressure instrumentation to leak/burst events or changes in the status of throttle valves.

(2) To develop systematic procedures for improved network analysis and calibration based on the available information and the sensitivity of the decision variables. Before solving the *inverse problem* the approach reduces its dimensionality to reveal a targeted starting point search space for subsequent optimisations.

(3) To further develop and expand the functionalities of the optimisation algorithm, in terms of its:
    (a) computational efficiency measured by number of simulations required to obtain optimal solution;
    (b) ability to precisely identify global optimum, i.e., to get as closely as possible (preferably exactly) to the global optimum;
    (c) ability to perform robust search in terms of method convergence and stability. This will include optimisation of: (i) throttle valve status and functionality to detect inflows at unknown open inlet boundaries, (ii) pipe roughness coefficient values, (iii) pressure reducing valve profile, (iv) variable speed pump profile and (v) demand pattern multiplier coefficients.

(4) To analyse the sensitivity of this methodology with respect to the accurate detection and localisation of leaks based on a number of different sensor configurations.

(5) To demonstrate, verify and evaluate the decision-support framework in improving the hydraulic model accuracy and for finding leaks on a number of representative and real-life case studies from the UK.

Finally, special attention is given in this thesis to the following aspects:

(1) To theoretical and practical benefits regarding the proposed novel approaches.

(2) To quantitative and qualitative comparisons of the proposed approaches to: (1) the existing theoretical approaches previously described in the relevant literature; (2) the practical trial-and-error approaches, based on engineering judgement and (3) the existing optimisation-based software applications used by the water industry.

1.4 Thesis Layout

This thesis is divided into six chapters including this introductory chapter.

In Chapter 2 a review of the relevant literature is provided. The review provides insight into *inverse problems* in WDN modelling. An introduction to the parameter

estimation theory and the formulation of optimisation problems is also given. Then, a wide overview of the model calibration procedure and the available techniques for the identification and localisation of leakages in the WDNs is presented. It also highlights their individual advantages and disadvantages through a critical analysis of the approaches used. The main emphasis is given in those approaches that make use of a WDN hydraulic model.

In Chapter 3 a systematic search space reduction technique is presented as part of the proposed decision-support framework to fulfil objective one. This is based on the general Inverse Problem Theory and Genetic Algorithm Optimisation, analysed in Chapter 2. The aim is to reduce the *inverse problem* dimensionality, which ultimately leads to a reduced search effort during an optimisation analysis. The decision-support framework is divided into two stages: (a) a Search Space Reduction stage, where the number of decision variables and the range of possible values is reduced; and (b) an Inverse Problem Solving stage, which considers the reduced set of decision variables in an optimisation analysis to solve the *inverse problem*.

In Chapter 4 two practical simulation-optimisation methods are proposed that make use of the search space reduction technique presented in Chapter 3. Utilizing one or both of these methods rely on the starting model for leak localisation, i.e., whether it is calibrated or not. The two methods are developed along the objectives two and three. When a calibrated model is available, then, leak localisation can be performed. In the opposite situation, the starting uncalibrated model has to be calibrated first before performing leak localisation. Based on the above situations, the following methods were developed: (1) A Leakage Inspection Method, which highlights the leakage area and makes pinpointing of leaks faster; and (2) A Calibration Method, which improves the WDN model accuracy so it can serve as the basis for use in (1).

In Chapter 5 the applicability of the techniques introduced in the prior chapters is demonstrated through their application in a number of semi-real and real case studies at a DMA level with a single source. The results of a number of analyses carried out on four case studies are presented to address objectives four and five.

Artificially generated and real data from pressure and flow devices are used to implement the methodology on semi-real and real case studies, respectively. All case studies have the same purpose, which is to ultimately detect and localise a single leak event that has happened in a DMA.

In Chapter 6 a summary is made. After that, relevant conclusions are drawn. This is followed by suggestions for future research work.

1.5 Related Publications

The author has published the research related with the thesis in multiple conferences and journals. A list of published works that derived from or influenced the thesis development are presented next:

- Sophocleous et al., 2015: Conference paper that raises issues and challenges required to advance WDN Modelling in practice. These are associated with data availability, the calibration approach and the accuracy criteria, which affect the leak detection and localisation accuracy. This work motivated the development of methods to improve the quality of WDN hydraulic models.
- Sophocleous et al., 2016: Conference paper that presents a new graph-theory-based technique to simplify the *inverse problem*, before the leak detection and localisation procedure. This work motivated the development of the search space reduction methodology.
- Sophocleous et al., 2017: Conference paper that presents a two-stage approach for solving the *inverse problem* for leak detection and localisation. The search space is reduced initially using an optimisation analysis, before solving the *inverse problem* in the second stage. This work inspired the use of optimisation analysis as part of the search space reduction methodology.
- Sophocleous et al., 2017b: Conference paper that presents a systematic approach for narrowing down the search area for the detection of leaks and unknown closed valves, considering noise in pressure measurements. The practical aspect of the search space reduction method in taking into

account the sensor configuration and the sensitivity of pressure instrumentation was driven by this work.

- Sophocleous et al., 2018: Conference paper that presents an overview and a real life application of a proposed leak detection and localisation methodology that simplifies the *inverse problem* before solving it. This work shaped the final version of the proposed *inverse problem* solving framework from this research.
- Sophocleous et al., (2018b – accepted): Journal paper that presents the proposed framework for solving the *inverse problem* for leak detection and localisation purposes. The framework is implemented on a semi-real and real cases.

CHAPTER 2

Inverse Problems in Water Distribution Network Modelling

2.1 Introduction

Physical theories allow us to make predictions, whereby given a complete description of a system we can predict its behaviour. To apply these theories, there must be a valid mathematical model of the system under study. The problem of predicting the behaviour of a specified system is called the solution to the *forward problem*. When the *forward problem* has been completely solved, there are unknown parameters in the mathematical model, representing the physical properties of the system. The goal of "Inverse Theory" is to determine the values of those parameters using available observations by means of data collecting devices, i.e., solving an *inverse problem* (Tarantola, 2005). The name *inverse problem* originates from the fact that one has to first know a *forward problem*, which describes the dependency of the given observations on the unknown parameters. Therefore the *forward problem* aims to determine the consequences from causes and the *inverse problem* the opposite.

A hydraulic model of the Water Distribution Network (WDN) is a mathematical description of the real system's hydraulic behaviour. During a hydraulic simulation analysis the physical properties of the system are considered to predict the hydraulic measurements of the WDN, such as pressure and flow, i.e., solving the *forward problem*. To ensure that the simulated behaviour matches the real WDN conditions as much as possible, hydraulic models are calibrated (Walski, 1983). This is an *inverse problem*, as actual measurements are taken from the field and are used to determine the values of the WDN model parameters. An accurate WDN hydraulic model can be used for a variety of applications, such as for finding leaks. Detection and localisation of leaks in the WDN by means of hydraulic modelling is also an *inverse problem*, which aims to identify parameter values

associated with the whereabouts and size of leaks. When a leak happens in a well-monitored WDN, it creates a unique “signature” on pressure and flow data, which can be used to automatically find its size and location. This brings multiple benefits of localising leaks and improving hydraulic model calibration, as by finding leaks the model becomes more accurate in simulating the real situation.

This chapter aims to identify the gaps in knowledge in the field of *inverse problems* in WDN modelling. These gaps, then, form the basis for the work done in the thesis. A literature review is presented, associated with model calibration and leak detection and localisation. Bearing this in mind, the chapter is organised as follows. After this introduction the theoretical framework of the *inverse problem* formulation and solution is firstly addressed in Section 2.2. In Section 2.3 the important role of optimisation techniques is addressed when solving *inverse problems*. Then, Section 2.4 and Section 2.5 analyse applications of the “Inverse Theory” in Model Calibration and in Leak Detection and Localisation, respectively. Finally, in the Section 2.6, a summary is provided and relevant conclusions are drawn.

2.2 The Inverse Problem

2.2.1 Background

In an *inverse problem* in WDN modelling a number of locations in the system are monitored and some network parameters are assumed known. On the other hand, some parameters are unknown (e.g., leaks, demands, pipe states, valve statuses), which affect the system’s hydraulics. These parameters cannot be determined explicitly by direct measurement, or there is no available data for them. The aim of the *inverse problem* is to determine the values of the unknown WDN model parameters that represent the real system behaviour. Hence, they are determined implicitly by comparing available observations (typically pressures and flows) with simulated outputs, as opposed to the *forward problem*, where these parameters are assumed known and are used to predict the system behaviour. This is demonstrated schematically in Figure 2.1.

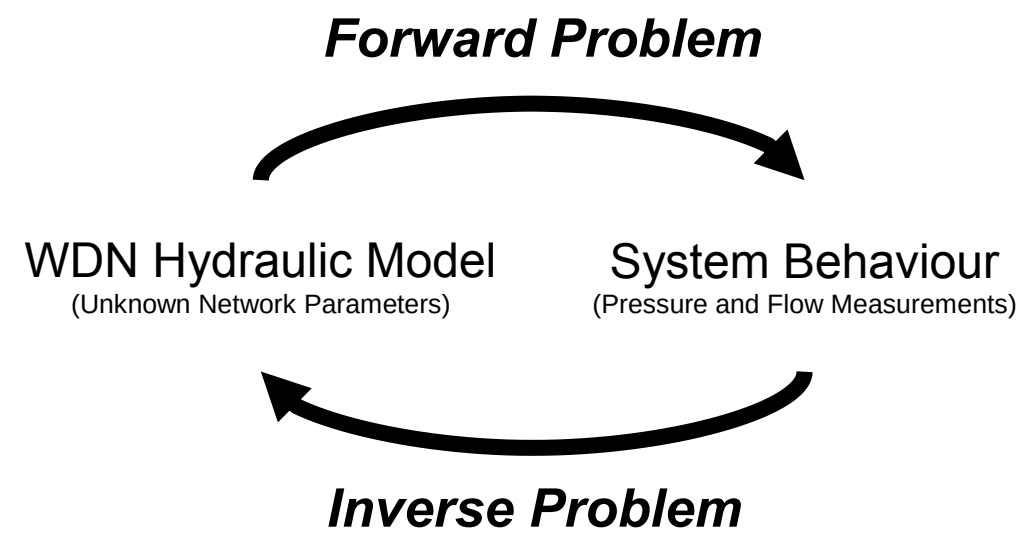

Figure 2.1. Forward vs Inverse Problem.

When measurements are available for every unknown parameter the *inverse problem* is characterised as *even-determined*. If the number of measurements is larger than the unknown parameters, then the problem is *over-determined*, in contrast to the *under-determined* case where there are more causal parameters (Menke, 2012). An *over-determined* case is the most desirable from the point of view of parameter determination as it gives more dependable results so that the unknown parameters can be determined. However, due to the non-linear nature of WDN models and the limited availability of field data relative to the number of parameters to be estimated, this results in an *under-determined* problem (Savic et al., 2009). A partly monitored system, however, can still give information. In addition, if there are fewer measurements than parameters, the causal parameters cannot be determined uniquely. Finally, a system can be *mixed-determined* where there are as many or more measurements and equations than unknowns, but still insufficient information to find a unique solution to the problem.

2.2.2 Problem Formulation

The general *inverse problem* (Tarantola, 2005) in WDN modelling (Pudar & Liggett, 1992) aims to determine the values for a set of parameters, x, called "decision variables". This is so that the discrepancies, between some set of measured values and the corresponding set of WDN model predicted (simulated) values, referred as "residual" or "error", are minimized. The ultimate purpose can

be for model calibration, leak detection and localisation, or any other problem that requires the analysis of the hydraulic model. In the general case, the residual consists of the following two basic parts: (1) modelling error, i.e., the discrepancy between true value (known to Mother Nature only) and the simulated value and; (2) measurement error, i.e., the discrepancy between the true value and the measured value. Both modelling and measurement error have a systematic and/or random component (Koppel & Vassiljev, 2012). An example of systematic type error is associated with uncalibrated measurement devices, e.g., an offset of readings. In contrast, the random component is usually a consequence of the imperfectness of the measuring device. The set of measurements or observations, y_o, collected from deployed sensors at certain locations, m, in the WDN typically record pressure and flow data. The device reading, y_o, comprises of two components, the true value, y_a, and the measurement error, ε, associated with the device accuracy range (systematic uncertainty), given by:

$$y_o = y_a + \varepsilon \tag{2.1}$$

The set of model predictions $y_s(x)$ that correspond to the observations are a function of the decision variable values, typically determined based on a criterion for the quality of solutions to the *inverse problem* (Shamir & Howard, 1968). This is represented by the residual errors, associated with the differences between measured and model predicted values, given by:

$$r = y_o - y_s(x) \tag{2.2}$$

In mathematical terms, the quality of a solution is expressed through an Objective Function (OF) and is associated with the values of the decision variables. The *inverse problem* is formulated as a constrained optimisation problem of weighted least-square type OF, given by:

Minimize: $f(x) = \sum_{m=1}^{M} w_m (r_m)^2$ (2.3)

Where: $f(x)$ is the OF value to be minimized;

w_m is the weighting vector corresponding to observations, y_o.

The weight vector, w_m, has two main related functions (Hill, 1998): (1) To reduce the influence of less accurate observations and increase the influence of those that are more accurate; and (2) To produce residuals that have the same units, i.e., that can be summed in Equation 2.3. With respect to the OF, several

alternatives exist, such as the weighted sum of absolute residuals. This method evenly spreads the influence over all observed datasets, as opposed to the squared residuals which places more emphases on bad data points because of squaring the difference values (Wu et al., 2011). However, the squared residuals OF demonstrated a better convergence of the gradient type method (Reddy et al., 1996). Conversely, as pointed by Vitkovsky & Simpson (1997), and Chen & Brdys (1995), a weighted sum of absolute residuals may be preferable in other cases.

Inverse problems become particularly interesting (and difficult) when the solution is required to satisfy hard (i.e., required) and/or soft (i.e., desirable) constraints, as in the case of model calibration and leak detection and localisation. These ensure that feasible, robust and stable solutions are produced, which can improve computational performance. However, naturally, sometimes adding constraints to an *inverse problem* will reduce the achievable performance. A feasible solution in WDN modelling is subject to two sets of constraints:

(1) Implicit System Constraints, defined by the relevant mass balance and energy/momentum equations, which are implicitly satisfied by using a hydraulic simulation model (Walski et al., 2004).

(2) Explicit Constraints, associated with maximum and minimum bounds for the value of each decision variable.

2.2.3 Well-Posed Vs Ill-Posed Inverse Problem

When searching for a solution to the *inverse problem*, the concern is not just finding a mathematically acceptable solution, as by their nature, such problems may usually have more than one solution. This characterises the problem as non-unique, or not well-posed mathematically. The essential conditions for a well-posed problem were introduced by Hadamard (1923) as follows (Figure 2.2):

(a) Solution Existence, i.e., there exists a solution for all admissible data.

(b) Solution Uniqueness, i.e., there is at most one solution to the problem.

(c) Solution Stability, i.e., the solution is robust against noise.

Typically none of these criteria is satisfied and almost always at least one is not satisfied. In such situation the problem is called as an *ill–posed inverse problem*.

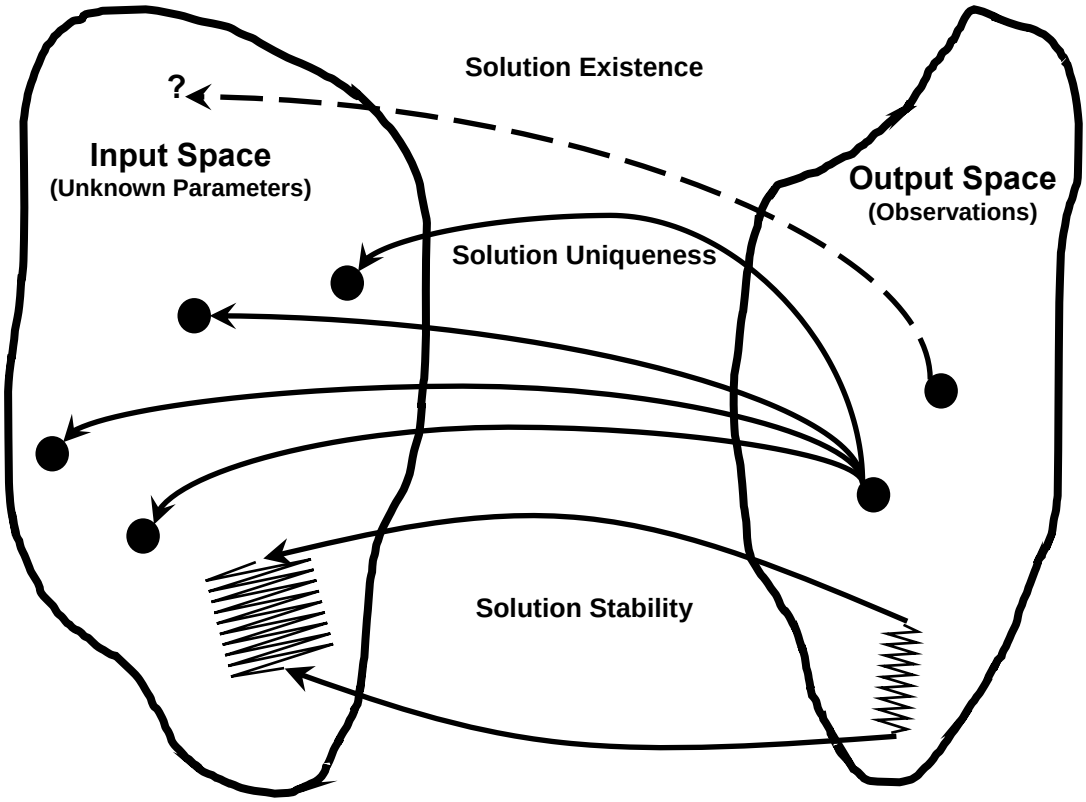

Figure 2.2. The conditions of a well-posed inverse problem.

There may be no solution that exactly fits the data (Yu, 1991). Therefore, before attempting to determine any parameters, the data that are associated with the model, need to be characterised (Parker, 1977). “Characterisation” concerns *solution existence* (Bilicz et al., 2010; Unser, 2016). It is of great importance to test the assumptions behind any mathematical model, as it contains simplifications and approximations, some of which may be hard to justify initially. A solution may fail to exist because the mathematical model of the system is approximate, or because of noisy data (e.g. corrupted, distorted, etc.). If the output space is defined as a set of solutions to the *forward problem*, the *existence* of a solution is clear. Nevertheless, *solution existence* is really a non-issue in realistic situations because the physical reality must be a solution.

If there is a solution for a given set of data, then, it is necessary to determine whether there is only one such solution, or there are many even for an infinite number of exact data points (Parker, 1977; Suzuki, 1983). The “Identifiability” problem is concerned whether there is enough data to determine the solution uniquely (Kool et al., 1987). Profound consequences follow if the solution is non-

unique as even perfect data (complete and exact) do not contain enough information to determine the unknown parameters. This is because different combinations of parameter values lead to similar observations. *Uniqueness* of a solution to the *inverse problem* is an important issue, however, it is often not easy to prove (Santamarina & Fratta, 2005). An important and thorny issue with problems that have non-unique solutions is that an estimated model may be significantly smoothed or otherwise biased relative to the true situation. One useful strategy to handle the non-uniqueness issue is to utilize a priori information as additional constraints (Bekey & Kogan, 2003). A more aggressive approach would be to use a Bayesian approach and incorporate prior knowledge probabilistically, such as the Tikhonov regularization (Tikhonov, 1963). Another fruitful approach is via search. Given a set of measurements and some defined constraints, optimisation algorithms can be used to search a hypothesis space of solutions to the problem. If uniqueness is not guaranteed by the given data, then either additional data have to be collected or the set of admissible solutions has to be restricted using a-priori information on the solution. In other words, a remedy against non-uniqueness can be a reformulation of the problem.

Mathematically, a problem is said to be stable if the solution depends continuously on the initial data (Bertero et al., 1980; Baumeister, 1987; Kabanikhin, 2008). This means that for all data sets lying close to a particular set of values, the solutions fall close to each other (Parker, 1977). A solution is often extremely unstable, whereby a small change in measurements can lead to an enormous change in the estimated set of unknown parameters. In this case, inevitable measurement and round-off errors can be amplified by an arbitrarily large factor and make a solution completely useless. It is possible to stabilize the inversion process by imposing additional constraints that bias the solution, a process that is generally referred to as regularization (Cheng & He, 2011). Regularization is frequently essential to produce a usable solution to an otherwise intractable ill–posed *inverse problem*, by limiting the solution space, i.e., replacing an ill-conditioned problem with a better conditioned problem (Tikhonov & Arsenin, 1977; Petrov & Sizikov, 2005).

2.2.4 Practical Considerations

Inverse problems in WDN modelling are often ill-posed, usually characterised by non-uniqueness and instability of the identified parameters (Yeh, 1986; Kool et al., 1987; Groetsch, 1993). Therefore, if possible, it would be valuable to know under what condition(s) the solution to the *inverse problem* is identifiable, unique and stable. System identifiability resolves whether the parameters of the system can be determined, i.e., whether different vectors of decision variables x may lead to (almost) the same vector of model predictions, $y_s(x)$, that are close to observations, y_o. As noted by Carrera & Neuman (1986), the problem of identifiability is closely related to problem of the uniqueness of the calibration problem solution, however, significant difference exists. Uniqueness, concerns exactly the opposite question, i.e., whether different values of parameter vectors may originate from the same vector of measurements and, thus, multiple parameter vectors correspond to similar values of OF. Identifiability condition is necessary but not sufficient for the uniqueness of the *inverse problem* solution. A sufficient condition for uniqueness is that the OF is convex within the domain of definition of the decision parameters (Carrera & Neuman, 1986). The information necessary to judge the problem of identifiability is available in Kapelan, 2002.

The problem of *solution stability* is also important, that deals with how small changes in the data disturb the corresponding solutions. It occurs when small observation errors, ε, lead to significant errors in the identified parameters x. Obviously, the problem of instability is closely related to the problem of identifiability. Parameters that are difficult to identify are those that are, typically, causing stability problems. Except from that, in practical *inverse problems* as in WDN modelling the available measurements are never exactly the same as in the mathematical formulation. There are several reasons for this (Walski et al., 2004; Santamarina & Fratta, 2005):

1. Pressure and flow devices have limited accuracy.
2. The *forward problem* theory is not necessarily completely correct. It may contain approximations.
3. There can be external disturbances in the measurement environment.
4. In numerical calculations, the real numbers are replaced by floating point numbers that are of finite accuracy.

Overall, the *inverse problem* condition can be improved (Kabanikhin, 2008; Santamarina & Fratta, 2005) by (a) reducing the dimensionality of the parameter vector, i.e., by reduced x; (b) considering alternative parameterisation schemes of the *inverse problem*, i.e., considering alternative vectors, x; (c) increasing the quantity of observations using additional field measurements.

2.2.5 Strategy for solving Inverse Problems

Successful *inverse problem* solving starts before data inversion. In fact, the first and most important step is to develop a detailed understanding of the WDN boundaries and underlying system hydraulics. Then one must set clear and realizable expectations and goals for the solution to the problem. The following steps provide a robust framework for the solution of *inverse problems* (Santamarina & Fratta, 2005): (1) Analysis of the problem; (2) Design of sensor placement; (3) Collection of high-quality data and pre-processing; (4) Selection of a hydraulic model of adequate detail; (5) Solving the *inverse problem* based on various methods; (6) Analysis of the problem solution.

Having an acute understanding of the problem is necessary to establish clear expectations for the problem solution. This involves knowledge of the: (a) underlying physical processes and constraints; (b) measurement and transducer-related issues; and (c) inherent difficulties in the *inverse problem*, such as non-linearity, the number of unknowns, the number of measurements and available information, which all relate to the ill-posedness issue.

The type, number, locations, accuracy range and monitoring frequency of the installed sensors in the field determines the viability of a solution (Aral et al., 2010). Therefore the sensor configuration placement plan is critical for collecting data of high quality (Walski, 1983). The sampling design procedure distributes devices and thus, measurements, to attain a proper coverage of the solution space (De Schaetzen, 2000; Ostfeld & Salomons, 2004; Ostfeld et al., 2008). The ultimate purpose is to collect data that, when used for the relevant *inverse problem*, will yield the optimum results (Kapelan et al., 2004). The sampling design addresses two critical aspects: (a) the distribution of measurements at optimal locations to attain a good coverage of the solution space, associated

specifically with the problem solved; and (b) the selection of instrumentation type and number to allow the collection of a good-quality and sufficient set of measurement. The optimal sampling design procedure depends on the specific problem that is investigated (e.g., model calibration, leak detection and localisation) and aims to determine the optimum type, number, locations and timeliness of observations at the least cost.

High-quality data are needed to improve the condition of the *inverse problem*. Data pre-processing permits identification and removal of outliers and provides valuable information to guide and stabilize the solutions, which can ultimately be used to generate a viable initial guess of the solution to the unknown parameters, x. A suitable strategy that permits diagnosing problems in the raw sensor data, associated with offsets or any other faults on devices is necessary. The simultaneous display of readings collected at neighbouring locations or time steps is particularly convenient to spot sudden changes in the system, to diagnose and remediate equipment failures, and to identify possible outliers. It is also important to be aware of how accurate each device is, which allows to differentiate the weight of information conveyed from each device. This can also help identify salient characteristics of the system. For instance, by performing simple computations and graphical display strategies insight is gained with respect to the measurements (noise level, outliers, and spatial coverage) and the solution characteristics (mean properties, spatial trends, presence of anomalies).

A hydraulic model of the WDN with an adequate level of detail must be considered that properly captures all essential aspects of the system. An inappropriate model represents the system inaccurately, adds error and hinders the inversion of a meaningful solution. In addition, the time required to compute the model is crucial if a massive forward simulation strategy is necessary to solve the problem, thus, computational demands should always be taken into account.

The problem should be solved based on various inversion methods, such as heuristics and the parametric representation of the problem, or less constrained representations of the problem and with a reduced number of unknowns. For repetitive problems, multiple forward simulations can be run and a library of

"solved cases" can be assembled as a scenario guide that can be used to identify an initial guess by data matching.

The physical and practical meaningfulness of the output solution to the problem must be analysed. The solution may be completely wrong even when the data are well justified and the residuals are small. Indeed, this is very likely in an ill-posed situation. Finally, the procedure that was followed to obtain measurements should be reanalysed to understand if all planned measurements were taken. The assumptions for the underlying physical processes should be reassessed in light of the results that were obtained and consider all available information. Furthermore, any systematic error propagation or accidental error magnification should be investigated and any correlation between parameters should be scrutinised. The solved *inverse problem* can convey unprecedented information, thus, in all cases, the hydraulics of the system should lead the way.

2.3 Optimisation and Inverse Problems

2.3.1 Optimisation Problem Formulation

The *inverse problem* can be solved by employing a search, or optimisation technique. Optimisation is the process of finding the conditions that minimize or maximize the value of a function which represents the effort required or the desired benefit. It is essentially the act of obtaining the best result under the given circumstances. With hard constraints in WDN analysis, optimisation becomes a “natural” way to proceed for solving *inverse problems*. An OF is used to measure how close the model outputs are to the observations, and a search is then conducted to find the parameter set which minimizes the OF value. Any optimisation problem is comprised of three basic ingredients (NEOS, 1996):

(1) **An Objective Function**, to be minimized, or maximized (Equation 2.3).
(2) **A set of Unknowns, or Decision Variables**, whereby their values are adjusted so that the OF value changes.
(3) **A set of Constraints** that allows the decision variables to adopt certain values or combinations of values, but excludes others so that feasible solutions are produced.

The optimisation problem aims to find the set of decision variable values, which minimize or maximize the OF while satisfying the constraints, formulated as:

Search for: $\vec{X} = (x_i); \qquad i = 1, \dots, DV$ (2.4)

Minimize: $F(\vec{X})$ (2.5)

Subject to: $\underline{x_i} \leq x_i \leq \overline{x_i}$ (2.6)

Where: $\vec{X}$ is the vector of decision variables to be identified (similar to x);
x_i is the i-th decision variable to be identified;
DV is the number of decision variables;
$F(\vec{X})$ is the OF (similar to Equation 2.3), usually formulated as a fitness function to be minimized, or maximized;
$\underline{x_i}$ is the minimum (lower) bound and $\overline{x_i}$ is the maximum (upper) bound for the i-th decision variable x_i;

Although, typically, optimisation problems have a single objective function, in other cases a problem can have multiple objective functions (Farmani et al. 2005). These are optimised simultaneously, leading to a set of equally efficient alternative solutions called the "Pareto-optimal set" (Marques et al., 2015; Tanyimboh & Seyoum, 2016). The different objectives are generally conflicting, resulting in non-dominated solutions, where the variables which improve one objective will cause worsening of at least one of the others. Problems with multiple objectives can be reformulated to single-objective problems by either forming a weighted combination of the different objectives or by replacing some of the objectives into constraints.

2.3.2 Search Space of Optimisation Problem

Optimisation searches for the best solution among a set of possible solutions. The space, or domain of all feasible solutions, i.e., the set of solutions among which the desired solution resides, is called search space (also state space). Any point in the search space represents one possible solution that is associated with a fitness value. The search process for finding the optimal solution to the optimisation problem works as follows:

1. Choose a “solution” from the search space and evaluate it. Define this as the current solution. In WDN modelling solution evaluation is performed by firstly solving the *forward problem* where the values $y_s(x)$ are calculated. Then, the residuals r are determined and used to find the OF (Eq. 2.3).
2. Transform the current solution to create and evaluate a new solution.
3. If the new solution is better than the current solution then exchange it with the current solution, otherwise discard the new solution.
4. Repeat steps 2 and 3 until no transformation in the given set improves the current solution.

The fitness landscape helps visualise the search space as a surface with peaks and troughs. The height of each point is analogous to the solution fitness value. The task of finding the best solution to the problem, i.e., the global optimum, is equivalent to finding the highest or lowest point, depending on the aim of the OF. The computational cost of a search process relies on the size and complexity of the search space. In addition, the time complexity of the optimisation algorithm (i.e., solution generation/transformation) and the hydraulic solver (i.e., solution simulation).

The complexity of the fitness landscape can severely affect the search process in reaching the global optimum as a result of: (a) local optima, (b) plateaux (flat local optima or shoulders), and/or (c) ridges (Figure 2.3). Depending on the starting point of the search process, the optimisation algorithm may converge prematurely to a local optimum, i.e., a state that is optimal within a limited part of the fitness landscape (Russel & Norvig, 2010). A similar stagnation in the search process can happen if a plateau in the search space landscape is reached, where the OF is constant in an area around it (Michalewicz & Fogel, 2004; Hoos & Stützle, 2004). In a plateau situation whose edges go downhill it is also called as a “flat local optimum”, whereas if there is an uphill edge, then it is called a “shoulder” (Figure 2.3).

Typically the solution space can be explored through Random search (Zabinsky, 2011), Gradient (Ruder, 2016) or Direct Methods (Lewis et al., 2000). A Random search algorithm refers to an algorithm that uses some kind of randomness or

probability and is useful for ill-conditioned global optimisation problems. In Gradient methods the gradient of the objective function is used to guide the direction of search. Finally, in Direct methods the objective function drives the search exploration. Gradient and Direct methods only perform well with unimodal functions, i.e., with a single optimal value, since they can be trapped in a local optimum with multi-modal functions. Random search and gradient methods are often combined together.

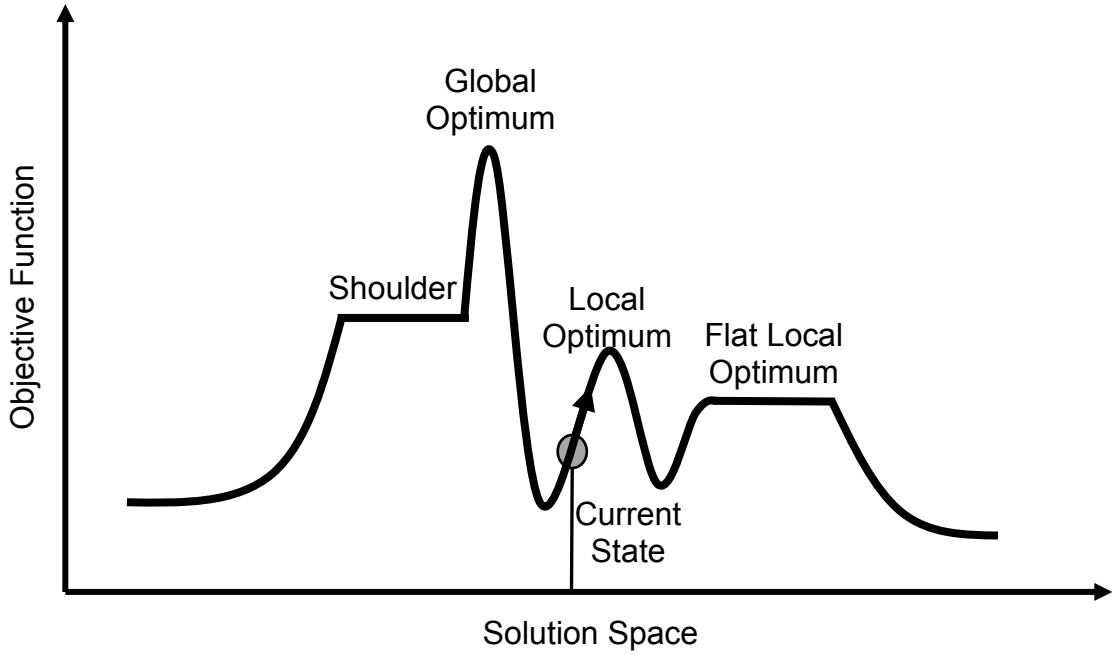

Figure 2.3. Example of Landscape in search space.

2.3.3 Genetic Algorithm Optimisation

In a WDN hydraulic model a large number of parameters cannot be easily optimised through standard non-linear function optimisation techniques. The search can often be fooled into declaring convergence far short of the true optimum. This occurs because of high dimensionality and irregularities contained in the fitness (OF) landscape (Figure 2.3) such as multiple optima, un-smoothness, discontinuity, elongated ridges, fiat plateaus, and so on. These difficulties, however, may be overcome to a large extent using Genetic Algorithms (GAs). These type of algorithms are a group of adaptive search methods, or metaheuristics, which may be used to solve various optimisation problems.

A GA belongs to a class of non-deterministic search algorithms based on artificial evolution, i.e., an artificial form of Darwin's evolution theory, in other words an evolutionary algorithm (Holland, 1975; De Jong, 1975; Goldberg, 1989; Bäck et al., 1997). GAs were introduced into the area of *inverse problem* solving for WDN model calibration by, Savic & Walters 1995. The method is inspired by biological evolution and the mechanics of real genetic processes and natural selection, through an artificial survival-of-the-fittest framework. They emulate nature's evolution using an artificial analogy of preferential survival, reproduction of the fittest members of the population, maintenance of a population with diverse members, inheritance of genetic material from parents, occasional mutation of genes, etc. A population of solutions is created and evolved among generations, whereby natural selection ensures that solutions with better fitness will propagate in the next populations.

A typical GA search process starts with the random generation of a population of individual chromosomes. Each chromosome represents a single possible solution to the optimisation problem and is associated with a fitness value according to the OF (Equation 2.5) and any introduced penalty functions (Camponogara & Talukdar, 1997). After the evaluation of the OF, a pair of individuals are selected from the population as the parents to reproduce the next generation of chromosomes, based on a probability proportional to their fitness. This imposes the survival-of-the-fittest mechanism on the candidate solutions. Parts of the parental solutions are recombined through a crossover operator, to produce new solutions, namely the child solutions.

To ensure that the solutions in a population do not become stuck at a non-optimal solution, a randomization element is introduced following recombination known as "mutation". Mutation introduces new features into the solutions to maintain diversity in the population and permits local search around a given solution. The mutation probability is generally kept low for steady convergence, as otherwise the solution finding process mimics a random search. Eventually, a new population of solutions is reproduced, and the fitness of each of them is evaluated. The least fit chromosomes are replaced when creating a new generation. An elitist strategy may be used to preserve the best chromosomes

within the next generation (Goldberg, 1989). The GA process typically terminates when a pre-specified number of generations is reached, or alternatively, when complete convergence has occurred. Besides convergence criteria, premature convergence is another issue, where the GA becomes trapped in a local optimum. A way to tackle premature convergence can be through fitness scaling, increasing the population size, adding non-random chromosomes in the initial population and increasing the mutation rate (South et al., 1993).

2.3.4 Optimisation Methods Comparison

GAs are often referred to as a global adaptive search technique best suited to solving large scale combinatorial optimisation problems which cannot be solved using conventional operational research methods. They differ from traditional optimisation methods (e.g., direct, gradient methods) in five significant points:

- They search parallel from a population of points. Therefore, it allows for efficient exploration of large, complex, multi-modal search spaces and provides the ability to avoid being trapped in local optimal solution like traditional methods, which search from a single point.
- They use probabilistic selection rules, not deterministic ones.
- They work on the “Chromosome”, which is encoded version of potential solutions’ parameters, rather the parameters themselves. Therefore, they can deal with a large number of parameters.
- They use fitness score, which is obtained from objective functions, without the need other derivative or auxiliary information. Thus, they can deal with non-smooth and noisy objective functions.
- They can be easily modified for solving different problems and can handle multiple objectives in parallel.

The GA and its many versions have been popular in academia and the industry mainly because of its intuitiveness, ease of implementation, and the ability to effectively solve highly nonlinear, mixed integer (discrete and continuous) optimisation problems that are typical of complex engineering systems (Hassan et al., 2005). The drawback of the GA is its expensive computational cost.

Particle Swarm Optimisation (PSO) is a metaheuristic search method whose mechanics are inspired by the swarming or collaborative behaviour of biological

populations (Kennedy & Eberhart, 1995). Like a GA, PSO is also a population-based search method that uses a combination of deterministic and probabilistic rules. Although PSO and GA share many similarities, it has no evolution operators (as crossover and mutation in GA) and, thus is easier to implement. Compared to GAs, the advantages of PSO are that there are few parameters to adjust and the execution and convergence time is inexpensive. However, PSO converges fast because it can be implemented without too many parameters and operators. The main disadvantage of PSO is its poor local search ability. The mutation and crossover operators will help GA to jump the discontinuity in the search space and lead to better exploration. On the other hand, PSO will get stuck in a disconnected component of the search space. An important distinction is that GA is a mixed-integer technique suitable for non-linear combinatorial problems, whereas PSO is a continuous technique that is very poorly suited to combinatorial problems. In addition, if the search space is discrete and is highly constrained and discontinuous, as in WDN modelling, the GA can outperform in the quality solutions, as opposed to unconstrained non-linear problems with continuous variables.

One of the latest evolutionary computational techniques apart from the PSO is the Differential Evolution (DE) algorithm (Storn & Price, 1995). DE is a stochastic direct search optimisation method that encodes solutions as vectors and uses operations such as vector addition, scalar multiplication and exchange of components to construct new solutions from the existing ones. The idea behind the DE method is that the difference between two vectors yields a difference vector which can be used with a scaling factor to traverse the search space. Like the GA, DE initiates with a random populations and allows each successive generation of solutions to ‘evolve’ from the previous generations and uses similar operators, i.e., crossover, mutation and selection. The DE differs from GA with respect to the mechanics of mutation, crossover and selection performed. GA relies on crossover while DE relies on mutation operation. In GA, the mutation takes place randomly, whereas in DE, it takes place by some rule. In DE, each variable’s value is represented by a real number and, thus, cannot deal with mixed-integer problems compared to the GA.

Simulated Annealing (SA) is another algorithm which is popular in metaheuristic optimisation (Metropolis et al., 1953; Kirkpatrick et al., 1983). SA is a single-solution-based random-search technique that belongs to a class of algorithms called probabilistic hill-climbing (Romeo & Sangiovanni-Vincentelli, 1991). It exploits an analogy between the way in which a metal cools and freezes into a minimum energy crystalline structure (the annealing process) and the search for a minimum in a more general system. SA's major advantage over other methods is an ability to avoid becoming trapped in local minima and exploitation of the search space as it searches the local solution space "around" its initial solution, which tends to find local improvements efficiently. The disadvantage against a GA is that SA is seeded randomly and so is not efficient in exploring large solution spaces. In addition, it modifies and improves as single solution, rather than multiple candidate solutions as in a GA.

2.3.5 Search Space Exploration Enhancement

Obviously, the nature of the search space landscape (Figure 2.3) dictates how an optimisation algorithm will perform. Except from that, solving optimisation problems in WDN modelling involves various constraints. The difficulties in solving constrained optimisation problems arise from the challenge of finding good feasible solutions. However, the cardinality of the set of feasible solutions is generally too intractably large and the cost evaluation of all feasible solutions is, thus, impossible in a reasonable time. The problem is much more challenging when the feasible space is very tiny compared to the search space. Solving such problems requires serious computational effort in finding the feasible space.

When the search space to be searched is large, it may often prove more efficient to first reduce it based on prior knowledge, and then to search that reduced space, instead of just searching the entire, initial search space. The gain will be increased if the search space will be searched several times, perhaps to find solutions characteristics. However, when reducing the search space, care must be taken not to eliminate the optimal solutions to the original problem, or more generally, the interesting solutions to this problem. Search space reduction restricts the original search space to some promising sub-space. The purpose of

search space reduction is to move some of the randomly generated initial poor solutions towards the feasible region, as demonstrated by Figure 2.4.

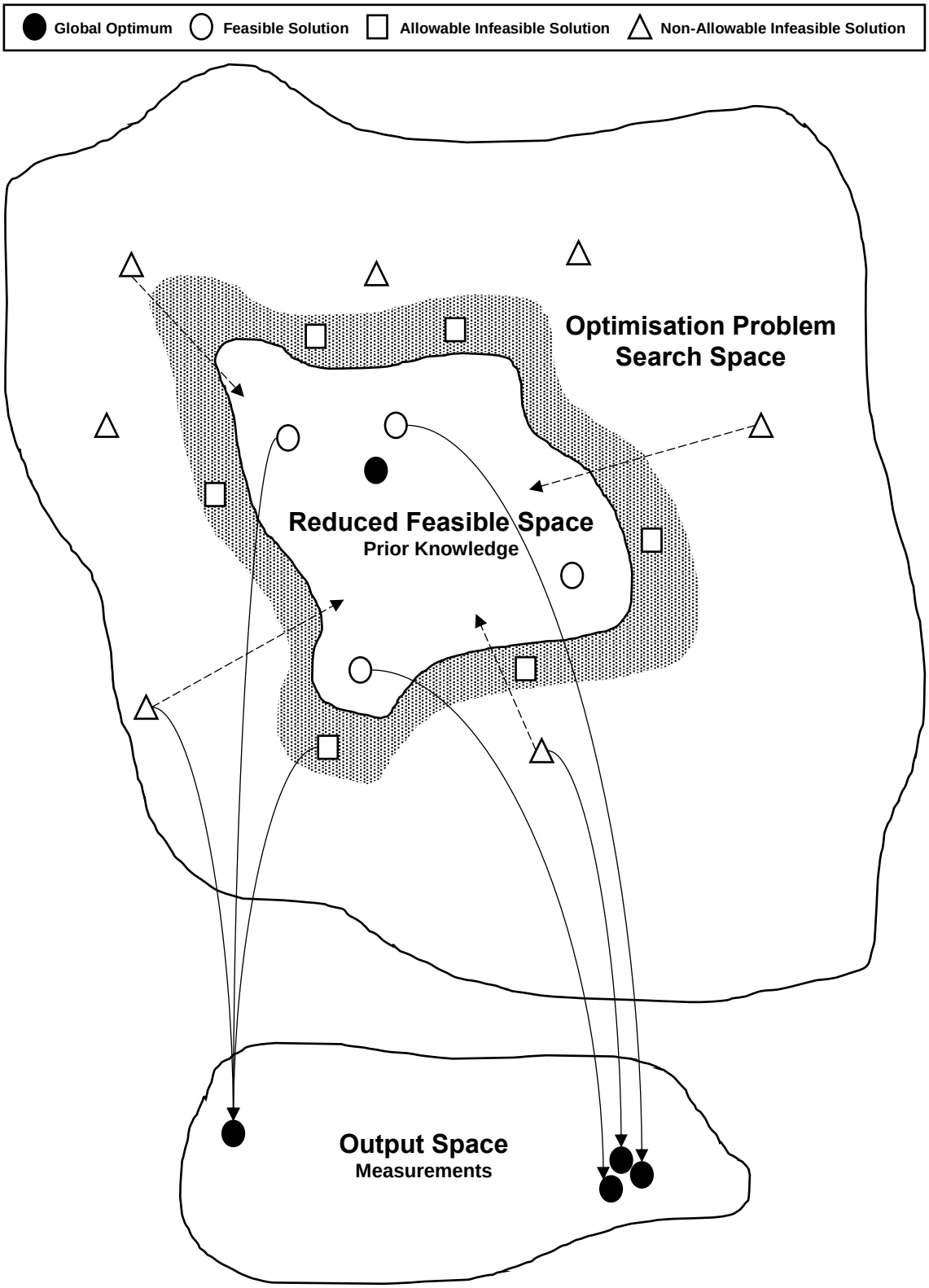

Figure 2.4. Search Space Reduction.

In population-based algorithms the quality of the initial population plays an important role on their performance. As the initial population is randomly generated to ensure diversity, it may cause delay (being over diversified) in reaching a reasonably good solution for tiny feasible space. If the initial population contains some good solutions the algorithms converge quickly. However, it is not expected that the random solutions should always be of good quality.

To enhance the performance of the algorithm in reaching the feasible space quickly the original search space can be reduced before starting the evolutionary process. Hence, the initial population is generated at a better starting point moving towards the feasible region of solutions. The solutions search for the global optimum from the feasible search space with some good infeasible solutions (here we are considering those solutions having less constraint violations). By applying search space reduction the randomly generated solutions are no longer random rather they have learnt a direction towards the feasible space which helps the algorithms to reach the feasible region faster and improve the solution quality.

2.4 Water Distribution Network Model Calibration

2.4.1 Introduction

A WDN hydraulic model is a mathematical description of the real system's hydraulic behaviour. Before it can be used it must be ensured that it predicts the behaviour of the WDN with a reasonable accuracy, i.e., it must be calibrated. Calibration is an *inverse problem* in WDN modelling, called the "calibration problem". The parameters describing the system are adjusted until the model-predicted performance reasonably agrees with the observed system performance, i.e., the field observations. Shamir & Howard (1977) stated that calibration "consists of determining the physical and operational characteristics of an existing system and determining the data, that when input to the computer model will yield realistic results". Walski (1983) proposed a more precise definition, "Calibration of a WDN model is a two-step process consisting of: (1) comparison of predicted pressures and flows with observed pressures and flows for known operating conditions (i.e., pump operation, tank levels, pressure

reducing valve settings); and (2) adjustment of the input data for the model to improve agreement between observed and predicted values.

A model is considered calibrated for a set of operating conditions and water uses if it can predict flows and pressures with “reasonable agreement”. The calibration process may include changing system demands, fine-tuning pipe roughness, altering pump operating characteristics, and adjusting other model attributes. All these adjustments affect simulation results and are necessary for the following reasons:

- **Confidence:** Calibration demonstrates the model’s capability to reproduce existing conditions, thereby increasing the engineer’s confidence that the model accurately predicts the system behaviour.
- **Understanding:** The calibration process provides excellent insight into the system behaviour and performance, as it can indicate which parameters and parameter values the model is most sensitive to.
- **Troubleshooting:** The ability to uncover system and data anomalies describing the WDN, such as incorrect pipe diameters, missing pipes, or closed valves.

On the other hand, model calibration challenges still remain (Savic et al., 2009) as there is no ideal WDN model, which can reproduce the real WDN conditions perfectly. Thus, the hydraulic model may not match the field observations. This can be because the cumulative effect of approximations, simplifications, uncertainties and errors can be so great that the model cannot distinguish between alternatives. According to AWWA (1999), calibration guidelines, various sources of errors, i.e., uncertainties exist which may influence the prediction accuracy of the calibrated WDN model, including: (a) Measurement readings; (b) Internal pipe roughness values; (c) System demands; (d) System map and System configuration e) Node elevations; (f) Tank levels; (g) Level of detail; (h) Geometric anomalies (e.g., crossing pipes, isolated valves, etc.); (i) Outdated pump characteristic curves, etc.

2.4.2 The Calibration Procedure

Calibration of hydraulic models is a routine component of the modelling process where field measurements are necessary. The process generally first requires a series of field tests during which pressures and flows are recorded at strategic locations in the system, normally continuously over one or more days. This is followed by a desk exercise during which adjustments are made to the parameter values used in modelling the system until a satisfactory match is obtained between modelled and observed values. Common calibration errors can be due to unsuspected cross connections, wrongly assumed throttle valve status, by-pass valves left open, valves between pressure zones left open or leaking and unexpectedly low demands from large consumers (Burrows, 1999).

If no satisfactory match is obtainable, further site checks are usually made to investigate for issues, such as leaks, incorrect valve positions, unrecorded connections etc. The calibration process is an iterative one. Also, note that it is necessary to calibrate WDN hydraulic models periodically to reflect physical changes in the WDN. Generally the calibration procedure consists of the following basic steps (Ormsbee & Lingireddy, 1997; Ormsbee, 1989; Walski, 1995): (1) Identification of the intended use of the model; (2) Determination of initial estimates of the model parameters; (3) Collection of calibration data; (4) Evaluation of the model results; (5) Macro-level calibration; (6) Sensitivity analysis; (7) Micro-level calibration.

2.4.3 Data Collection Process

The data collection process for calibration is carried out through an activity commonly known as field testing. Flows in selected pipes and pressures at selected hydrants are recorded simultaneously. This field data is also typically supplemented with measurements of large consumer demands and detailed monitoring of control valves, pumps and reservoirs. A well-designed field test provides a good coverage of the WDN in terms of the type, number and locations of the devices, allowing to gather high-quality calibration data (Aral et al., 2010). A typical field test involves the following tasks (WRc, 1989):

(1) A field test plan or sampling design is defined, i.e., a plan of the proposed field test monitoring locations. The period of testing and the frequency of recording are also defined prior to the field test.
(2) Zone boundary valves, bypass valves to meters and control valves are then checked to ensure that they are all shut.
(3) Pressures and flows are continuously recorded simultaneously at the monitoring locations for the specified field testing period.

Flow monitoring points are placed at all the import and export points of the system, inlets and outlets of water storage and at supplies to major non-domestic consumers. Pressure monitoring points are usually placed on important hydrants distributed throughout the WDN, which are modelled as junction nodes. For each monitored location, elevation measurements are taken which are used to calculate total head values. Those located on dead-end branches are rarely monitored due to the small and intermittent head losses occurring in those branches. Additional pressure monitoring points may be required where system problems are known to exist. However, the exact locations of those sensors in a WDN are often selected by subjective judgement and therefore do not ensure optimal or near optimal data collection for model calibration. Fire flow tests may also be undertaken. Such field testing methods stresses the WDN by opening one or more hydrants to increase the demand artificially. Although fire flow tests are more expensive than normal field tests, they produce generally more accurate estimates. The greater the flow carried in the pipes, the greater the head loss that is produced to accurately estimate the roughness values or detect unknown closed valves. It is often necessary to close some of the valves in order to isolate the tested pipes.

2.4.4 Calibration Approaches

Once the field test has been completed the model calibration process begins. Numerous calibration procedures have been developed since the 1970s, which can be generally grouped into three categories (Savic et al., 2009; Tabesh et al., 2011): (1) iterative (trial-and-error) procedure models; (2) explicit models (or hydraulic simulation models); (3) implicit models (or optimisation models). Iterative calibration models are based on some, specifically developed, trial-and-

error procedure (Rahal et al., 1980; Walski, 1983; Bhave, 1988). The current approach used by most of the water companies is still the traditional manual method of trial-and-error, whereby hydraulic tables are used to assist the modeller in estimating the roughness values depending on the pipe type.

The modellers first estimate the parameter values and then run the model to compare model predictions with observations. If model predictions are not close enough to observed data, modellers return to adjust parameters and run the model again to obtain new predictions. Such procedures generally have to repeat many times, leading to time-consuming tasks. Historically, the main benefit from development of iterative procedures is in the establishment of some fundamental principles and guidelines regarding WDN model calibration. These principles were utilised later on in the development of more sophisticated explicit and implicit calibration models of improved efficiency.

Explicit calibration models are based on solving an extended set of steady-state mass-balance and energy equations (Ormsbee & Wood, 1986; Boulos & Wood, 1990; Boulos & Ormsbee, 1991; Ferreri et al., 1994). This extended set consists of initial equations (describing a steady-state network model) and a number of additional equations derived from available head and flow measurements (one additional equation per measurement). The extended set of equations is solved explicitly, usually by the Newton-Raphson method. Obviously, the number of unknown calibration parameters is limited by the number of available measurements, due to the need to translate the calibration problem to be *even-determined*. Another limitation is that measurements are assumed to be 100% accurate. In addition, there is no way to quantify uncertainty in the estimated calibration parameters. Finally, these methods require considerable mathematical expertise and sophisticated solution tools, giving them only historical significance and no apparent influence on the current practice of model calibration.

Implicit calibration refers to problems which are formulated as optimisation problems (Shamir, 1974; Ormsbee, 1989; Pudar & Liggett, 1992; Savic & Walters, 1995; Reddy et al., 1996; Greco & Del Giudice, 1999; Tucciarelli et al.,

1999; Todini, 1999; De Schaetzen, 2000; Kapelan, 2002; Wu et al., 2002; Kapelan et al., 2007; Walski et al., 2008; Wu & Clark, 2009; Koppel & Vassiljev, 2009; Alvisi & Franchini, 2010; Cheng & He, 2011; Wu & Walski, 2012; Ostfeld et al., 2012; Dini & Tabesh, 2014; Puust & Vassiljev, 2014; Vassiljev et al., 2015; Do et al., 2016; Xie et al., 2017). The *inverse problem* is solved using an optimisation technique coupled with a hydraulic solver, whereby using an OF the differences between the observations and simulated outputs are minimized (Figure 2.5). A summary of the key studies solving the calibration problem using optimisation techniques is given in Table 2.1.

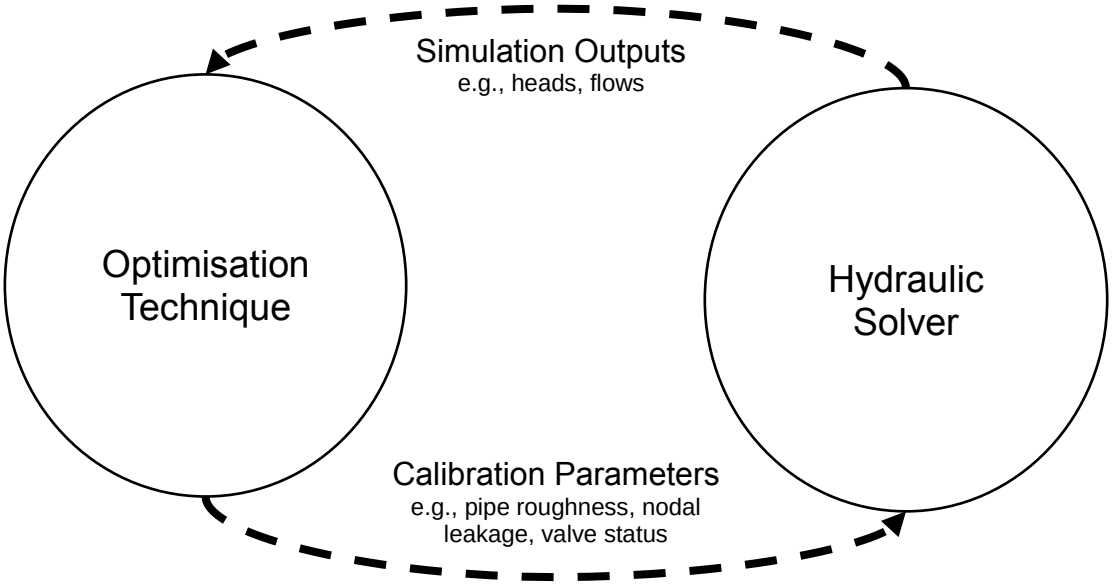

Figure 2.5. The Implicit Calibration Procedure (Savic et al., 2009).

The set of constraints associated with this problem are implicit hydraulic constraints, known initial conditions (device statuses and tank levels), and boundary conditions (see Section 2.2.2). Rather than explicitly incorporating the equations of conservation of mass and energy into the optimisation process, later approaches have simply called out to a standard hydraulic simulation program to evaluate the hydraulics of the solution (Ormsbee, 1989; Lansey & Basnet, 1991). The most used network hydraulic modelling software that is freely available is EPANET 2 (Rossman, 2000). The stochastic search procedure of a GA is typically used, where the optimisation problem is formulated as:

Search for: $\vec{X} = (f_i, m_{j,t}, s_{k,t}); \quad i = 1, \dots, NI; \; j = 1, \dots, NJ; \; k = 1, \dots, NK$ (2.7)

Minimize: $F(\vec{X})$ (2.8)

Subject to: $\underline{f_i} \leq f_i \leq \overline{f_i}$ (2.9) $\underline{m_{j,t}} \leq m_{j,t} \leq \overline{m_{j,t}}$ (2.10) $s_{k,t} \in \{0,1\}$ (2.11)

Where: $\vec{X}$ represents a set of decision variables;

$\underline{f_i}$ and $\overline{f_i}$ are the lower and upper limits of roughness coefficient for the i-th pipe or pipe group;

$\underline{m_{j,t}}$ and $\overline{m_{j,t}}$ are the lower and upper limits for the demand adjustment multiplier for the j-th junction or junction group at time step t;

$s_{k,t}$ is the status of k-th throttle valve at time step t;

$F(\vec{X})$ is the OF that measures the goodness-of-fit between the field observed values and the model simulated values (Equation 2.5).

For an effective calibration process, when using implicit models it is critical to understand which parameters can be calibrated with confidence, as well as the acceptable level of discretization of calibration parameters. In addition, to establish how the quality of a solution is evaluated, through an appropriate OF. Finally, an acceptable level of agreement between the observations and model predictions is necessary.

2.4.5 Model Performance Criteria and Validation

Regardless of which calibration approach is adopted, a realistic model of the WDN should meet some performance criteria, i.e., a minimum required level of accuracy for calibrated model predictions (Hydraulic Research, 1983). In the United Kingdom, certain performance criteria have been established by the Water Research Centre (WRc, 1989) that are still in place and designers strive to meet these standards (Table 2.2). On the other hand, there are no explicit calibration guidelines in the United States.

Table 2.1. Summary of Implicit Calibration Procedures.

Author	Calibration Parameters[1]	Optimisation Method[2]	Objective Function[3]	Test Example Number of: Pipes/Nodes/Pressure Sensors
Shamir, 1974	Any	GRG	General Form	40/25/4
Ormsbee, 1989	FC, ND	Extended Complex Method of Box	WSAE	40/16/6-18
Pudar and Liggett, 1992	LLC	LM	WSSE	11/7/1-3; 35/20/9
Savic and Walters 1995	FC	GA	WSSE	242/197/237
Reddy et al., 1996	RC, ND	GN	WSSE	11/7/5; 73/57/7; 94/87/3-6
Greco et al., 1999	FC	LINDO, GINO	SSE	16/14/16
Tucciarelli et al., 1999	TDEM, LP, VR	SA, QN	WSSE	28/18/8
Todini et al., 1999	FC	KF	SSE	12/09/2012
De Schaetzen, 2000	FC, PD, ND	GA	RMSE	21/13/4-21; 34/16/5; 519/447/171
Kapelan, 2002	FC	Hybrid GA	(W)SSE	497/451/10
Wu et. al., 2002	FC, ND, LS	fmGA	WSAE, WSSE, ME	46/32/2
Kapelan, 2006	FC	SCEM-UA	SSE	40/16/4; 497/451/10
Walski, 2008	FC	GA	WSSE	12/14/5
Wu and Clark, 2009	FC, ND, LS	fmGA	WSAE, WSSE, ME	-/-/20
Koppel and Vassiljev, 2009	FC	LM	SSE	2436/-/25
Alvisi and Franchini 2010	FC, ND, LS	SCE-UA	SSE	443/399/-
Tabesh and Khanal, 2011	FC, ND, PD	GA	WSSE	16/12/2004
Cheng and He, 2011	ND	ODS	SSE	12/11/3/2; 855/492/22/1
Ostfeld et al., 2012	FC, VS, PS	SCE-UA, LM, GA, HS, CLONALG, DDS	SSE	443/399/-
Dini and Tabesh, 2014	FC, ND	ACO	SSE	8/7/4; 192/165/27
Puust and Vassiljev, 2014	FC	LM, GA	SSE	2436/-/18
Vassiljev et al., 2015	FC	LM	WSSE	2436/-/18

Abbreviations:

1) FC - Friction Coefficient, RC - Generalised Resistance Coefficient, PD - Pipe Diameter, LS - Link Status, VS - Valve Setting, VR - Valve Resistance, PS - Pump Speed, ND - Nodal Demand, TDEM - Total Demand, LLC - Lumped Leak Coefficient, LP - Leak Parameters,
2) GA - Genetic Algorithm, fmGA - fast messy GA, LM - Levenberg-Marquardt, GRG - General Reduced Gradient, GN - Gauss-Newton, SA - Simulated Annealling, QN - Quasi Newton, KF - Kalman Filter, SCEM - Shuffled Complex Evolution Metropolis, HS - Harmony Search, CLONALG - Clonal Selection, DDS - Dynamically Dimensioned Search
3) SSE - Sum of Squared Errors, WSSE - Weighted SSE, RMSE - Root Mean Squared Error, SAE - Sum of Absolute Errors, WSAE - Weighted SAE, ME - Maximum Error,

Table 2.2. Calibration criteria for flow and pressure.

Flow Criteria
(1) Modelled flows should be within ±5% of the measured flows when the measured flow is more than 10% of the total demand.
(2) Modelled flows should be within ±10% of the measured flows when the measured flow is less than 10% of the total demand.
Pressure Criteria
(1) Modelled pressures should be within ±0.5m or ±5% of the maximum head loss across the system for 85% of field test measurements.
(2) Modelled pressures should be within ±0.75m or ±7.5% of the maximum head loss across the system, for 95% of field test measurements.
(3) Modelled pressures should be within ±2m or ±15% of the maximum head loss across the system, for 100% of field test measurements.

Nevertheless, many modellers agree that the level of effort required to calibrate a WDN hydraulic model and the desired accuracy level relies on its intended use (Ormsbee & Lingireddy, 1997; Cesario et al., 1996; Walski, 1995). Each application of a model is unique, and thus it is impossible to derive a single set of performance guidelines to evaluate calibration. The true test of model calibration is that the end user feels comfortable when using it to assist in decision making. After a model is calibrated to match a given set of test data, the modeller can gain confidence in the model and/or identify its shortcomings by validating it with test data from a period not used during model calibration, i.e., obtained under different conditions. Validation is an essential processes for quantifying and building confidence in WDN modelling and is concerned with quantifying the model accuracy by comparing simulated outputs to unseen test data, i.e., observations that were not used for the calibration process. Although it is desirable to validate every model, most utilities do not have the time or money required to perform a thorough verification of the entire system.

2.5 Leak Detection and Localisation in Water Distribution Networks

2.5.1 Introduction

Leaks often remain undiscovered, regularly resulting in large losses of water and revenue for WDN operators. With time, their impact in the WDN grows and can

provoke catastrophic bursts, which can alter the system's operation causing devastating consequences for the customers and the utility. In many WDNs, losses from leaks are estimated to account for up to 30% of the total volume of extracted water (Puust et al., 2010). This comprises a vital amount in a world struggling to satisfy water demands due to growing population and climate change. Leakage events differ as distinction can be made between bursts, leaks and 'background' leaks. Bursts are normally reported by consumers, as they became visible early following the pipe failure. Yet, the leaks that do not cause water to come up at the surface, can remain undiscovered for a long time and ultimately have similar or larger consequences as large bursts. Finding leaks at an early stage can save water and prevents small leaks turning into bursts, which is important to a water company for economic, environmental and reputational reasons.

A range of strategies are adopted by water utilities to fulfil the purpose of controlling and reducing water losses (Goodwin, 1980; Water Authorities Association UK, 1980; Pearson & Trow, 2005). The traditional approach to leakage control is a passive one, whereby water companies take actions for a leak only when it becomes visible, or the event is discovered through 'ad-hoc' investigations following customer complaints, referred as "reported leaks". This strategy is clearly simple and is a lower cost option for operating expenditures on resources, facilities and equipment. However, it is time consuming as it relies heavily on customer contacts and surveillance campaigns by the utility. Ultimately, this results in a higher leakage levels, due to the overall volume of water lost from undiscovered/unreported leaks, which run for a long period of time (Puust et al., 2010). Furthermore, a reactive approach presents inevitable problems for the customers (Ramos et al., 2001), depending on how fast a leak is repaired.

Effective and efficient leakage management strategies involve water companies acting more proactively into targeted WDN areas, so that leakage can be controlled or reduced (Pearson & Trow, 2005). A typical integrated leakage management model is associated with activities such as Speed and Quality of Repairs, Pressure Management, Active Leakage Control and Infrastructure

Management (TWGWW, 1980; Lambert, 2001; Thornton, 2003). On the other hand, leak-free WDNs are not a realisable technical or economic objective, and a low level of leakage cannot be avoided, even in the best operated systems where water suppliers pay a lot of attention to leakage management. However, such activities aim to eliminate all the potentially recoverable water losses and reach the Unavoidable Annual Real Losses as closely as possible (Lambert & Lalonde, 2005). The extent of use of all such proactive approaches should be economic so according to the Economic Level of Leakage.

2.5.2 The Leak Detection and Localisation Procedure

Leakage can be minimized by shortening the detection, location and repair times. Historically, finding leaks has been challenging because even a substantial event can potentially show no manifest signs (Casillas Ponce et al., 2014). A wide range of leak detection and localisation techniques exists (Puust et al., 2010; Hamilton & Charalambous, 2013), however, there is no universally agreed methodology for finding leaks with the number of techniques currently used by practitioners being limited.

The term “leak detection” is used to explain the discovery, or “narrowing down” of a leak to a particular section of a pipe network (Kapelan et al., 2003; Mounce & Boxall, 2010; Palau et al., 2012; Romano et al., 2014). It does not give any information about its precise location, but only the leak size. Automatic leak detection requires pressure and flow field measurements. The smaller the monitored area, the easier it is to detect a leak automatically. Consequently, WDNs are often divided into smaller areas, called ‘District Metered Areas’ (DMAs), which are easier to monitor and control (Moors et al., 2018). It is normal practice to monitor the DMA for detecting leakages by making water balances between the expected and actual water use. In addition, DMAs are useful to identify any unexpected increase of flow during the night, i.e., when consumption is low and hence the majority of the flow recorded represents leakage (Farley et al., 2008). Still, leakage detection is not easy, due to unpredictable variations in consumer demands, measurement noise and seasonal effects.

To reduce the search area for finding a leak within a DMA, hydrostatic leak detection methods, such as step-testing or pressure-testing is often applied, by redirecting flows through systematic valve openings and closures, or making pressure adjustments, respectively. These methods identify the presence of leakage by subdividing the WDN during the period of Minimum Night Flow (MNF), while observing the corresponding reduction in flow or pressure. While these techniques can reduce the search area, their main disadvantage is the shutting-off supply to isolated parts of the network. This may be ineffective in networks with few isolation valves. They are also manually intensive and relatively expensive techniques as they require several steps over multiple nights because they provide information to a relatively small part of the network. In addition, they are not suitable for all DMAs (Puust et al., 2010) as they may interrupt water supply, cause back-siphonage or infiltration of groundwater, or create a burst, which are major reputational concerns for water utilities when adopting such approaches for finding leaks.

Minimizing the search area is only part of the problem after a leak has been detected, hence, to find its exact location a leak localisation technique is necessary. “Leak localisation” (Romano et al., 2013) refers to an activity that identifies and prioritises the areas of leakage to make pinpointing of the exact position of leaks easier (Pilcher et al., 2007). Leakage detection and localisation techniques (Farley & Trow, 2003; Colombo et al., 2009) divide into externally-based and internally-based techniques (ADEC, 1999). These are described in more detail in Sections 2.5.3 and 2.5.4.

2.5.3 Externally-Based Methods

Most frequently leak localisation is performed using externally-based methods, which use data from sensors installed or exploring outside of the pipe to pinpoint a leak. They are very accurate for finding leaks on pipelines, however, they take a long time to find a leak in a large search area (Li et al., 2015). Acoustic devices such as listening rods, stethoscopes, ground microphones, noise loggers and leak noise correlators are widely used in practice (Clark, 2012; Hartley, 2009). These devices rely on sound and vibration signals induced by leaks impacting

the soil from pipelines under pressure. The two traditional types of instruments used for finding leaks are the listening stick (stethoscope) and an electronic amplifier and detector. Technicians use their skill and experience to find the exact location of leaks in DMAs, often with the consultation of historical records from previous leak/burst locations (Pelletier et al., 2003).

The effectiveness of acoustic methods has been successfully demonstrated for both metallic and plastic pipes (Fuchs & Riehle, 1991; Hunaidi & Chu, 1999). However, conventional acoustic equipment can be unreliable for “quiet” leaks in non-metallic (e.g., plastic) pipes and large-diameter pipes, as the leak noise is strongly attenuated (Hunaidi, 2000; Mutikanga et al., 2013). Acoustic leak detection with correlation, i.e., leak noise correlators, have been developed in the market in the last 20 years to become the most common method of pinpointing leaks. They can accurately locate leaks with low flow rates between 0.05 and 3.5 m^3/h to within 1 metre in most pipe sizes (Guttermann, 2014; Echologics, 2015; Sewerin, 2015).

The disadvantages of acoustic technology is that the effectiveness of such methods depends on the operator’s experience, size of the leak(s), knowledge of exact pipe location and are limited to smaller size diameters of low depth. In addition, unwanted interference noise from traffic wind, water use etc., the leak shape, pipe material (e.g., acoustic dumping from plastic pipes) or the varying pressure conditions in the WDN may limit/change sound propagation (Seaford, 1994). For larger diameter pipelines (>300mm), in-service pipeline methods using acoustic sensors have been developed that are also able to provide information on condition of the pipeline necessary for strategic asset management (Mergelas & Henrich, 2005; Wu et al., 2011; Ong & Rodil, 2012). However, they require depressurizing mains to enable removing the sensor from inside the pipeline and may not be suitable for mains with bends and in locations that are not easily accessible by vehicles (Hamilton et al., 2012).

Recent times have seen the introduction of some new and innovative techniques (Li et al., 2015), based on non-acoustic methods such as ground motion sensors, ground-penetrating radar (Hugenschmidt & Kalogeropoulos, 2009), infrared

Table 2.3. Main characteristics of externally-based techniques.

Technique	Ability[1]	Cost[2]	Search Effort[2]	Advantages	Disadvantages
Patrolled Noise Listening	D/L/P	M	VH	• Easy to use equipment	• Relies on the technician's experience • Affected by the technician's fatigue and surrounding noise • Depends on the survey's degree of detail • Unreliable for non-metallic and large diameter pipes
Noise Logging	D/L	H	H	• Beneficial in WDN parts that have high surrounding noise during the day • Enable examining the sensor's response remotely	• Limited leak detection effectiveness Unreliable for non-metallic and large diameter pipes
Leak Noise Correlation	D/L/P	VH	H	• Enable suppressing the non-leak noise • Do not depend on the technician's experience	• Require extensive training • Unable to deal with multiple leaks • Leaks in branched pipeline segments require particular attention • Unreliable for non-metallic and large diameter pipes
Ground Penetration Sensing	D/L	H	H	• Reliable for non-metallic and large diameter pipelines	• Depends on the technician's experience • Impractical in highly conductive clay and saturated soils • Affected by the ambient conditions such as thermal noise and relative humidity (i.e., difficult to use in urban areas)
Infrared Thermography Sensing	D/L	H	VH	• Can also detect and locate erosion voids and poor backfill	• Depend on the technician's experience • Affected by the ambient conditions such as thermal noise and relative humidity (i.e., difficult to use in urban areas)
Inline Pipeline Inspection	D/P	H	VH	• Can also be used for condition assessment	• Affected by the presence of WDN elements (e.g., valves) and corrosion • Applicable to larger pipelines only • Entail contact with the inner surface of a pipeline
Tracer Gas Sensing	D/P	VH	H	• Able to deal with multiple leaks in branched pipeline segments Reliable for non-metallic and large diameter pipelines	• Impractical for routine surveys • Require isolation and dewatering of the inspected WDN section • Require expert contractors to carry out the work • Depend on the weather conditions and wind direction

Abbreviations: 1) D – Detection, L – Localisation, P – Pinpointing; 2) L - Low; M - Medium; H - High; VH - Very High.

thermography and pipeline inspection gauge have been found promising (Fanner et al., 2007; Hunaidi, 2000). In some rare instances tracer gases, using nontoxic odour gases or some electrically detectable gases, that are water-insoluble and lighter than air (e.g. helium, hydrogen, SF_6), can be injected into isolated sections of water pipelines, along with a gas detector to improve the efficiency of surveillance (Field & Ratcliffe, 1978; Hunaidi, 2000; Farley & Trow, 2003). However, such method is costly and its effectiveness strongly depends on the weather conditions and wind direction (Black, 1992; Furness & Reet, 1998). In general, the abovementioned techniques are time consuming and costly, labour intensive, often imprecise and invasive. The main characteristics of externally-based methods are summarized in Table 2.3. Automatic detection through the analysis of hydraulic measurements from permanently installed sensors is a more cost-effective method which can provide a rapid response to the on-set of a burst or leak event (Ye & Fenner, 2011).

2.5.4 Internally-Based Methods

2.5.4.1 Non-Mathematical Modelling

Internally-based methods use field sensors to monitor internal pipeline parameters, such as pressure and flow, and infer the position of leaks using mathematical and non-mathematical methods. Non-Mathematical Modelling methods use Artificial Intelligence techniques which need historical sensor data for training (Li et al., 2015). This approach is typically to construct a data-driven model by extracting the good information from a large amount of data samples. For example, many systems now automate the monitoring of night flows based on data-driven models (Armon et al., 2011), saving time and errors arising from manual interpretation.

The methods include: (1) traditional methods based on experience learning theory, such as the artificial neural network (Caputo & Pelagagge, 2002; Caputo & Pelagagge, 2003; Shinozuka et al., 2005; Feng & Zhang, 2006; Aksela et al., 2009; Mounce & Boxall, 2010; Tao et al., 2014; Romano et al., 2014) and the Bayes identification method (Poulakis et al., 2003; Rougier, 2005; Puust et al., 2010; Kapelan et al., 2007; Romano et al., 2009; Romano et al., 2011; Zhou et

al., 2011; Barandouzi et al., 2012; Hutton et al., 2014; Costanzo et al., 2014) and (2) machine learning methods based on the statistical learning theory (Vapnik & Kotz, 1982), such as the support vector machine (Mashford et al., 2009; Mounce et al., 2011; Candelieri et al., 2014; Mamo et al., 2014; Zhang et al., 2016).

The drawback of these techniques is that the required large amounts of training data are not always available (Li et al., 2015; Tao et al., 2014). When the size of the training samples is small, the performance of these methods will usually be affected. In addition, the applicability of these methodologies to real WDNs may be limited under certain circumstances, such as the hypotheses made which are usually impossible to meet due to systematic modelling errors and the limited experience and prior knowledge of the user.

2.5.4.2 Mathematical Modelling

2.5.4.2.1 Leakage Modelling

Mathematical Modelling-based techniques exploit information provided on the WDN flow conditions and by field measurements to make a diagnosis related to model performance. They use a hydraulic model of the WDN and compare simulated results with field data (Li et al., 2015) through: (a) steady state analysis (involving steady state, extended period simulation and real-time modelling) and (b) transient (unsteady) state analysis. The conventional network hydraulic solvers analyse WDNs based on the assumption that nodal demands are fixed and independent of network pressures commonly referred to as demand-driven analysis (DDA). These assumptions are increasingly being challenged and new modelling techniques and algorithms are emerging (Giustolisi et al., 2008; Wu et al., 2010; Mahmoud et al., 2017). DDA is only appropriate when WDNs are simulated under normal conditions with adequate pressures which in practice is not always the case, e.g., during pipe failure. Leakage is often implicitly included in nodal demands during design of WDNs which is also not realistic. Leakage is a type of Pressure Driven Demand (PDD) and must be explicitly considered in order to simulate hydraulic characteristics.

This realisation motivated researchers to develop methods for realistic leakage modelling (Germanopoulos, 1985; Vela et al., 1991). Several researchers have assessed leakage using WDN hydraulic simulations that fully incorporate leakage as PDD (Almandoz et al., 2005; Burrows et al., 2003; Giustolisi et al., 2008; Tabesh et al., 2009; Tucciarelli et al., 1999). As an alternative, in EPANET 2 (Rossman, 2000) a leak event can be simulated within a WDN by an orifice flow based on emitter hydraulics, as a means to represent leak flow discharge. Leakage along a pipe is allocated as a specified emitter coefficient to the connected nodes in a hydraulic model. The emitter nodes allow leakage to be modelled using appropriate pressure dependent outflow relationships, given as:

$$Q_i = c_i[p_i]^a \tag{2.12}$$

Where: Q_i is the leakage flow rate (l/s) at node i;
c_i is an emitter coefficient depicting the leak orifice area at node i;
p_i is pressure (m) at node i;
a is a pressure exponent commonly set to 0.5 under ideal conditions, in order to simplify the problem. However, a can vary from 0.5 to 2.5 with the pipe material and leak shape and may impact leak localisation (Lambert, 2002).

Equation 2.12 indicates that a positive emitter coefficient will result in leakage demand at a node. When its value is greater than zero at a node, that node is referred to as a leakage node or leakage hotspot indicating that leaks may exist on the pipes connected with the node. Applying a leak/burst event in this way has been done to match an increase in flow at the inlet. However, the volume of water lost from a leak will vary depending on the pressure in the system. As a result the volume lost from the leak will vary over the day as pressure changes.

2.5.4.2.2 Inverse Analysis Methods

The literature on model-based leak detection and localisation methods for WDNs focuses on how to prioritize areas for leak surveys and facilitate pinpointing of leaks. Such research started with the seminal paper of (Pudar & Liggett, 1992),

which formulates the leak detection and localisation problem as a least-squares *inverse problem* of parameter identification. In an *inverse problem* the system characteristics and the demands are known, except from some quantities (e.g. leaks, unaccounted nodal outflows), which are causal parameters for the system's hydraulics. If there are sufficient field observations these unknown parameters, i.e. the leak size and location(s), can be determined. However, due to the non-linear nature of WDN models and the limited availability of field data relative to the number of parameters to be estimated, this results in an underdetermined problem (Savic et al., 2009).

A variety of methods have been applied to solve for the *inverse problem*. Many of these techniques are based on pressure transient analysis in pipes (Liggett & Chen, 1995; Vıtkovsky et al., 2000; Wang et al., 2001; Covas, 2003; Kapelan et al., 2004; Taghvaei et al., 2006; Colombo et al., 2009). After the appearance of a leak, the pressure wave can accurately be localised by the use of sensors with high sampling intervals. However, these methods have been mainly used on a single, grounded pipeline due to the high effect of the exact system configuration, which is often unknown in real WDNs, the uncertainty on results and the unwillingness to generate water hammer in the system. This results in a costly and time consuming process that is computationally demanding, labour-intensive and requires large amounts of high-frequency data (Li et al., 2015; Puust et al., 2010).

Non-transient model-based leakage localisation techniques have been also developed recently (Wu & Sage, 2006; Wu et al., 2010; Farley et al., 2013; Goulet et al., 2013; Moser et al., 2018). These approaches analyse the difference between field observations and modelled outputs from leaky scenarios and signal an area experiencing leakage by assigning emitter flows as a way to simulate leaks. The *inverse problem* is solved by employing an optimisation technique (Figure 2.3), such as Genetic Algorithms (Savic & Walters, 1995; Kapelan, 2002), Particle Swarm Optimisation (Sreepathi & Mahinthakumar, 2013; Debiasi et al., 2014), Simulated Annealing (Sousa et al., 2015; Ribeiro et al., 2015), or other methods (Berglund et al., 2017; Kun et al. 2018), coupled with a hydraulic model to simulate the hydraulic responses for the possible leakage solutions

corresponding to the size and location of leaks. The general procedure requires collecting field data so that the objective function can be quantified for each solution. In order to identify leaks, the optimisation model needs to be formulated to optimise the node emitter coefficients (c_i in Equation 2.12) so that the differences between the observed and simulated responses (pressures and flows) are minimized. In the general formulation the optimisation is formulated as:

Search for: $\vec{X} = (q_l), \quad l = 1, \dots, NLeak$ (2.13)

Minimize: $F(\vec{X})$ (2.14)

Where: q_l is the l-th leak to be identified;

$NLeak$ is the number of leaks to be identified;

$F(\vec{X})$ is the objective function (Equation 2.5).

Wu & Sage (2006) combined leakage detection with hydraulic model calibration as a parameter identification problem with the same objective function and used a GA to solve the problem. The leakage emitter coefficients for each candidate leak location node were optimised and the approach showed success in identifying leaks in a small case problem and a real district metered area. However, in a real system, it would require the optimisation of hundreds of emitter coefficients to include all possible leak locations. Based on the same rationale the leak detection and localisation was defined as an *inverse problem* with the task of determining nodal demands and/or demand patterns (Wu et al., 2002; Cheng & He, 2011; Di Nardo et al., 2015; Sanz & Pérez, 2014; Sanz et al., 2016; Kun et al., 2015; Do et al., 2016). During the calibration process leakage was treated as nodal demand. By comparing the calibrated parameters with their historical values, the approach can identify if changes in these parameters are caused by leakage. However, an accurate estimation of consumer demands is necessary, because of their impact on pressure variation, unless the leak size is large relative to the inlet (Moors et al., 2018). To date, model-based methodologies have not reached the maturity required for mainstream adoption by water utilities.

The *inverse problem* is often ill-posed, characterised by the non-uniqueness of the identified parameters. Even in well-monitored systems multiple combinations of decision variable values could produce equally fit solutions, but inaccurate leak localisation. This is a consequence of small impact caused by a leak on the pressure and flow data due to a small velocity change in the pipe experiencing leakage or sub-optimal sensor placement. Currently, there is no optimisation algorithm that can efficiently and effectively solve a non-linear inverse problem with thousands of decision variables. A solution to this problem can be to reduce the search space without losing optimum solutions.

Wu (2009) and Wu et al. (2010) used inverse analysis, i.e., formulated an *inverse problem*, to detect and localise leaks. The developed method reduced the problem dimensionality by specifying the maximum number of possible leaks within a WDN. This is based on the fact that it is the nature of leakage that the sizable and restorable leaks can often be associated with a small number of candidate locations. Thus, they reformulated the problem such that it identifies a given number of leakage nodes and corresponding emitter coefficients. This was done by fixing the number of decision parameters, according to the number of leaks being identified, whereby the emitter locations and the corresponding emitter coefficients were optimised simultaneously. For each possible leak two decision variables were defined, one associated with the discrete location and one for the continuous emitter coefficient value, formulated as:

Search for: $\vec{X} = (LN_i^n, c_i^n);$ $LN_i^n \in J;$ $c_i^n \in K;$ (2.15)

$n = 1, \dots, NLeak;$ $i = 1, \dots, NIndex$

Minimize: $F(\vec{X})$ (2.16)

Subject to: $\underline{c^n} \leq c_i^n \leq \overline{c^n}$ (2.17) $P > 0$ (2.18) $NLdup^n = 0$ (2.19)

Where: LN_i^n is the index for node i for the possible leak n;

c_i^n is the emitter coefficient (equivalent to c_i in Equation 2.12), respectively, for node i for the possible leak n;

J is the set of potential leak locations for any possible leak;

K is the range of possible values for any possible leak;

$NLeak$ is the number of possible leaks to be identified;

$NIndex$ is the number of the candidate nodes for any possible leak;

$\underline{c_i^n}$ and $\overline{c^n}$ are the minimum and maximum emitter coefficients, respectively for the node i for the possible leak n along with a possible value of 0 that can be selected to indicate a no-leak case; P is the pressure at any WDN node and $NLdup^n$ is the number of the duplicate nodes that are identified as leakage emitters in the same solution. The constraints for pressure and the number of duplicate nodes are handled by using a penalty function and are necessary for the GA to search for good and realistic solutions. This leads to avoidance of solutions that may cause negative pressures in the system or those where a node is identified as a location of multiple leaks if $n > 1$.

Even though this method reduces the search space by assigning the leakage location a decision variable with a user specified maximum number of possible leaks within a system, it does not narrow down the number of candidate leak locations nor the range of flow values prior to conducting the leak localisation. Furthermore, the reduction process is severely subjective based on engineering judgement. Thus, it is easy for an optimisation algorithm to converge prematurely in a large search space, which affects the accuracy of the leak localisation.

Another optimisation-based approach, called step-by-step elimination method, was proposed in Nasirian et al. (2013) to calibrate the WDN hydraulic model and detect leaks. The approach was tested on hypothetical and laboratory networks, where nodes that were not reported as leaks during the optimisation process were eliminated. Although the search domain was reduced, the process requires iterations to converge to a reduced list of candidate locations before the final optimisation analysis. In a real WDN model this would be too time-consuming. Moreover, none of the abovementioned methodologies ensured that the optimum solution remained in the search space after the reduction. Therefore, if the specified number of possible leaks is lower than the true WDN state, or the true leak location is eliminated due to premature convergence during the first iterations, the solution would miss the optimum. A summary of all approaches that solved an *inverse problem* to find leaks in the WDN, is given in Table 2.4.

Table 2.4. Summary of Leak Detection and Localisation Procedures based on Inverse Analysis.

Author	Method[1]	Leakage Parameters[2]	Optimisation Method[3]	Objective Function[4]	Test Example Number of: Pipes/Nodes/Pressure Sensors
Pudar and Liggett, 1992	S	LOC	LM	WSSE	11/7/3 ; 35/20/10
Ligget and Chen, 1994	T	LOC	LM	SSE	11/7/3
Vitkovsky et al., 2000	T	LOC	GA	SAE	11/7/6
Kapelan et. Al., 2004	T	LOC	LM, GA	SSE	11/7/2
Wu and Sage, 2006	S	NDM	GA	SSE, SAE, ME	21/27/5; -/-/28
Wu et. al., 2010	S	LOC, LLI	GA	SSE, SAE, ME	11/7/1-3; -/-/28
Cheng and He, 2011	S	ND	ODS	SSE	12/11/3; 855/492/22
Nasirian et al., 2013	S	NDM	GA	WSSE	9/6/2; 50/30/7
Debiasi et. Al., 2014	S	LOC	GA, PSO, DE, HMO, ACO	SSE, SAE, ME	34/23/4
Sanz et al., 2015	S	LOC	-	SSE	3455/3377/5
Sousa et. Al., 2015	S	PLI,LF	SA	SAE	129/100/10
Ribeiro et. Al., 2015	S	PLI,LF	SA	SAE	111/100/10-20
Di Nardo et al., 2015	S	LOC	GA	SSE	-
Do et al., 2016	S	NDM	GA	WSSE	9/6/-; 12/11/3
Berglund et al., 2017	S	LOC	LP	SAE	34/31/31; 117/92/59; 28331/26986/45
Kun et al., 2018	S	ND	GN	SAE	62/43/3

Abbreviations:

1) S – Steady, T – Transient.
2) LOC – Leak Orifice Coefficient, LLI – Leak Location Index, ND – Nodal Demand, NDM – Nodal Demand Multiplier, PLI – Pipe Location Index, LF – Leak Flow.
3) LM – Levenberg-Marquardt, GA – Genetic Algorithm, PSO – Particle Swarm Optimisation, DE – Differential Evolution, HMO – Honeybee Mating Optimisation, ACO – Ant Colony Optimisation, SA – Simulated Annealing, LP – Linear Programming, GN – Gauss-Newton, ODS – One Dimensional Search.
4) SSE – Sum of Squared Errors, WSSE – Weighted SSE, SAE – Sum of Absolute Errors, ME – Maximum Error.

2.5.4.2.3 Residual Analysis Methods

Residual Analysis techniques are model-based leakage localisation techniques that exploit information provided by field measurements to make a diagnosis related to model performance, whereby a leak is identified based on a sensitivity analysis (Perez et al., 2011; Farley et al., 2012; Casillas Ponce et al., 2012; Casillas Ponce et al., 2014; Casillas et al., 2015; Sala & Kołakowski, 2014; Sarrate et al., 2014; Ferrandez-Gamot et al., 2015; Blesa et al., 2016; Soldevila et al., 2016; Soldevila et al., 2017; Gamboa-Medina & Reis, 2017). They seek for pressure anomalies between field measurements and simulated values from a hydraulic model in a 'leak-free situation' above a certain threshold. It depends on analysis of the residuals using a threshold that considers the modelling noise and uncertainty. When residuals exceed their threshold, an abnormal event can be determined by comparing against the leak sensitivity matrix to discover which of the possible leaks is present. Although this approach can work under ideal conditions, it has been shown that its performance decreases due to the nodal demand uncertainty and noise in the measurements.

For this reason, Pérez et al. (2009) have recommended to perform leak localisation during night hours when there is less consumption and hence less noise in the pressure data. This approach has been applied to a real network in Barcelona by Pérez et al. (2011, 2014) to identify the most probable location of the leak, while Meseguer et al. (2014) developed a decision support tool for online leak detection and localisation. A serious limitation for the practical applicability of this approach is that accurate estimates of the spatial distribution of customers' water demand are necessary, because of its influence on pressure variations and therefore also leak localisation performance (Cugueró-Escofet et al., 2015; Meseguer et al., 2014; Mirats-Tur et al., 2014; Perez et al., 2011; Sanz & Pérez, 2014; Sanz et al., 2016). Such detailed demand data are often unavailable. In addition, this method has only been demonstrated for small networks and for simplified scenarios such as single leak cases and zonal leak localisation with no regard to magnitude. Rosich & Puig (2013) and Rosich et al. (2015), generated a new class of structured residuals that allows to detect leaks in an efficient way, but it may be time-consuming due to the numerical algorithm used to compute the residuals. The methodology of Perez et al. (2011) has been improved in

Casillas Ponce et al. (2014, 2015), who introduced a time horizon analysis, called the Leak Signature Space, that associates a specific signature to each leak location being minimally affected by leak magnitude. In this case, the efficiency is improved despite the leak magnitude uncertainty still affecting the performance. Moreover, the resolution of the sensors strongly affects the performance of the leak location.

Goulet et al. (2013) proposed a leakage detection methodology based on an error domain model falsification whereby model instances (parameter sets) for which the pressure residuals were larger than the maximum plausible error were falsified. The method was able to find large leaks of 100l/min in a WDN with 295 pipes using 14 flow sensor devices. The detected flow is not adequate for full-scale applications, as practitioners are interested in detecting much smaller leaks with that level of monitoring. In addition in real WDNs such flow data availability is scarce. Moser et al. (2015) presented a methodology for simplification of WDN models combined with error-domain model falsification for leak detection. Using the same WDN as in Goulet et al. (2013), the computational time was reduced to 20% of the initial model, but the leak diagnostic performance reduced. Moser et al. (2018), applied the error-domain model falsification approach in a larger WDN demonstrating potential for practical use with less flow sensors (i.e., 10 devices). However, this work was only able to find a leak of large magnitude, i.e., 200l/min.

Soldevila et al. (2016), presented a new approach for online leak localisation in WDNs that can only be used once the leak has been detected. The approach relies on the computation of pressure residuals that are analysed through a Bayesian Classifier. This online scheme relies on a previous offline work, in which the network model is obtained and the classifier is trained with data generated in extensive simulations of the network. Considering the dynamic operation of WDNs and the always unpredictable appearance of leaks, the time required for such analyses may limit their practical use. Finally, Salguero et al. (2018), located leaks based on a simplified calculation of the sensitivity of the elements of the network, combining the knowledge of the specific average behaviour of each element following an increase in demand at any point of the network, and the information provided by the pressure sensors. However, with no model reduction

a real application would be computationally intensive. A summary of the main approaches that used forward analysis of the hydraulic model to find leaks, is given in Table 2.5. In addition the main characteristics of the presented internally-based methods are compared in Table 2.6.

Table 2.5. Summary of Leak Detection and Localisation Procedures based on Forward Analysis.

Author	Leakage Parameters[1]	Leak Diagnosis Method[2]	Test Example Number of: Pipes/Nodes/Pressure Sensors
Perez et al., 2011	ND	RM	-/1600/8; -/260/5; -/2132/10
Farley et al., 2013	LOC	RM	-/260-1091/3
Goulet et al., 2013	ND	MF	295/263/-
Perez et al., 2014	LOC	RM	3442/3377/6
Ferrandez-Gamot et al., 2015	ND	RM	34/31/3
Casillas et al., 2015	LOC	LSS	239/197/2-5
Soldevila et al., 2016	ND	RM	34/31/2
Moors et al., 2018	ND	RM	3243/3218/13
Moser et al., 2018	LOC	MF	295/263/-; 904/900/-
Salguero et al., 2018	ND	RM	58/38/5

Abbreviations:
1) LOC - Leak Orifice Coefficient, ND - Nodal Demand.
2) RM - Residuals Matrix, MF - Model Falsification, LSS - Leak Signature Space.

2.6 Concluding Remarks

In this chapter, the theoretical background of Inverse Theory and literature review associated with its applications on WDN model calibration and leak localisation was presented. Solving the *inverse problem* in WDN modelling is usually a very complex task that requires field measurements that are used to determine the unknown causal parameters, relevant to model calibration and leak localisation purposes. Rather than using some trial-and-error approach, an implicit type procedure should be used to solve it, where the problem (i.e., model calibration or leak detection and localisation) is formulated as an optimisation problem driven by the weighted least square type objective subject to a set of implicit and explicit type constraints (see section 2.2.2 for details). Therefore the unknown causal parameters can be determined effectively and efficiently. However, for a well posed problem to be formulated a solution must exist, be unique and be stable to changes on data.

Table 2.6. Main characteristics of internally-based techniques.

Technique	Ability[1]	Cost[2]	Search Effort[2]	Advantages	Disadvantages
Water Balance	D	VH	VH	• Enable identifying DMAs with the highest leakage levels • Does not require any field work • Enable identifying issues that prevent a water company from achieving high water delivery efficiency	• Rely on the accurate estimation of several night flow consumption components • Low sensitivity to leakages • Require 'out-of-hours' work • Carried out sporadically
Step/Pressure Testing	D/L	H	H	• Can narrow down the location of a leak within a DMA section/part	• Rely on the accurate estimation of several night flow consumption components • The shutting down of various DMA subsections can cause backsiphonage and the risk of infiltration of ground water • Create supply interruption inconvenience to the customers • Require 'out-of-hours' work
Artificial Intelligence	D/L	L	H	• Can effectively enable online monitoring of pipelines • Do not require high frequency measurements • Only rely on empirical observation of the WDN section behaviour • Provide an efficient and automated means for the analysis of large volumes of imperfect data • Can adapt to the changes in the WDN section's operating conditions	• Require large amounts of data samples • Allow the discovery of a leak in a particular DMA without giving any information about its more precise location
Mathematical Modelling	D/L	M	M	• Can narrow down the location of a leak within a DMA section/part • Does not require high frequency measurements • Able to deal with multiple leaks • Can be used automatically	• Rely on an accurate hydraulic model • Require accurate measurements at multiple locations • Involve manual and resource intensive processes for data logging, collection, and transfer to the point of use • Often require long computation times

Abbreviations: 1) D – Detection, L – Localisation, P – Pinpointing; 2) L - Low; M - Medium; H - High; VH - Very High

The issues associated with the ill-posedness of *inverse problems* were addressed and a strategy for solving them was presented. Subsequently, the efficient and powerful global search method of Genetic Algorithms was addressed, which are based on the genetic process of biological organisms and are capable of dealing with the non-linear *inverse problems*. However, in practical problems non-uniqueness and stability of solutions is a serious issues, due to the approximations of in physical theories and inadequate quantity and/or quality of observed information combined with the errors usually contained within.

The literature review on model calibration and leak detection and localisation presented various approaches to solve such *inverse problems* that are mainly costly, time consuming and labour-intensive. The following gap in knowledge has been identified, that forms the basis for the work done in this thesis:

"Both problems can be mathematically formulated as an optimisation problem, however, in a large search space it would be computationally too demanding to allow its effective use in practical applications. Several improvements can be made to condition an ill-posed problem, such as to reduce the search space, without losing optimum solutions. This allows the optimisation analysis to begin at a better starting point, while the computations required to explore the search space are minimized. However, none of the abovementioned techniques has tried to solve the inverse problem in by reducing the original problem."

Having recognised the need to improve the condition of the *inverse problem* in the next Chapter a novel search space reduction technique of the optimisation problem is developed and presented using prior information of the system and available data.

CHAPTER 3

Search Space Reduction of Inverse Problems in Water Distribution Network Modelling

3.1 Introduction

An optimisation analysis in WDN modelling requires many simulations of the *forward problem* which computations can be costly. However, by definition an *inverse problem* is an ill-posed problem. Even in well-monitored systems multiple combinations of decision variable values could produce equally fit solutions, but inaccurate results. This is a consequence of small or local impact caused by an anomaly in the WDN (e.g., a leak) on the pressure and flow data. Such hydraulic impact is due to an insignificant velocity change in the pipe experiencing it, or sub-optimal sensor placement. Except from that, the size of the WDN hydraulic models has significantly increased, which makes the inverse analysis much more complex process, due to the large dimensionality of the problem.

Currently, there is no optimisation algorithm that can efficiently and effectively solve a non-linear *inverse problem* with thousands of decision variables (Nyarko et al., 2014; Beiranvand et al., 2017; Song & Chen, 2018). The use of global optimisation algorithms, such as the Genetic Algorithm (GA), is hampered mainly by two facts: (a) The high computation time needed to solve a hydraulic simulation; and (b) The large, multi-dimensional parameter space to be explored to identify best parameter values. Therefore, global optimisation methods become intractable for high dimensional problems. A solution to this challenge can be to reduce the search space without losing the optimum solution, in order to make the *inverse problem* less ill-conditioned. Adopting the parameter set that

results in well-posed conditions also reduces the number of dimensions in which the *inverse problem* is going to be solved.

In Chapter 3 a search space reduction technique is presented that is developed based on the general Inverse Problem Theory (Tarantola, 2005) and Genetic Algorithm Optimisation, analysed in Chapter 2. The technique is part of the decision-support framework proposed in this thesis for solving the *inverse problem* in WDN modelling. The aim is to reduce the *inverse problem* dimensionality, which ultimately leads to a reduced search effort during an optimisation analysis. Bearing in mind the above, this chapter is organised as follows. After this introduction, an overview of the developed decision-support framework is presented in Section 3.2. The Sections 3.3, 3.4 and 3.5 provide insight on how the *inverse problem* size is reduced in three steps, using the developed search space reduction technique. Then, Section 3.6 provides the theoretical background on how the reduced *inverse problem* is solved. Finally, in Section 3.7 the summary and conclusions from this chapter are outlined.

3.2 Decision-Support Framework Overview

A model-based approach for reducing and solving the *inverse problem* is developed based on systematic search space reduction. The proposed decision-support framework is divided into two stages, shown schematically in Figure 3.1:

(a) a Search Space Reduction (SSR) stage, where the number of decision variables and the range of possible values is reduced using the developed search space reduction technique; and

(b) an Inverse Problem Solving (IPS) stage, which considers the reduced set of decision variables in an optimisation analysis to solve the *inverse problem*.

During the SSR stage, the number of decision variables in set, x, of the original *inverse problem* and the range of possible values is narrowed down prior to solving it, while trying to ensure that the optimum solution is not lost. The reduced set of decision variables, x_{SSR}, and values is, then, used as part of a simulation-optimisation framework, for solving the *inverse problem* in the IPS stage.

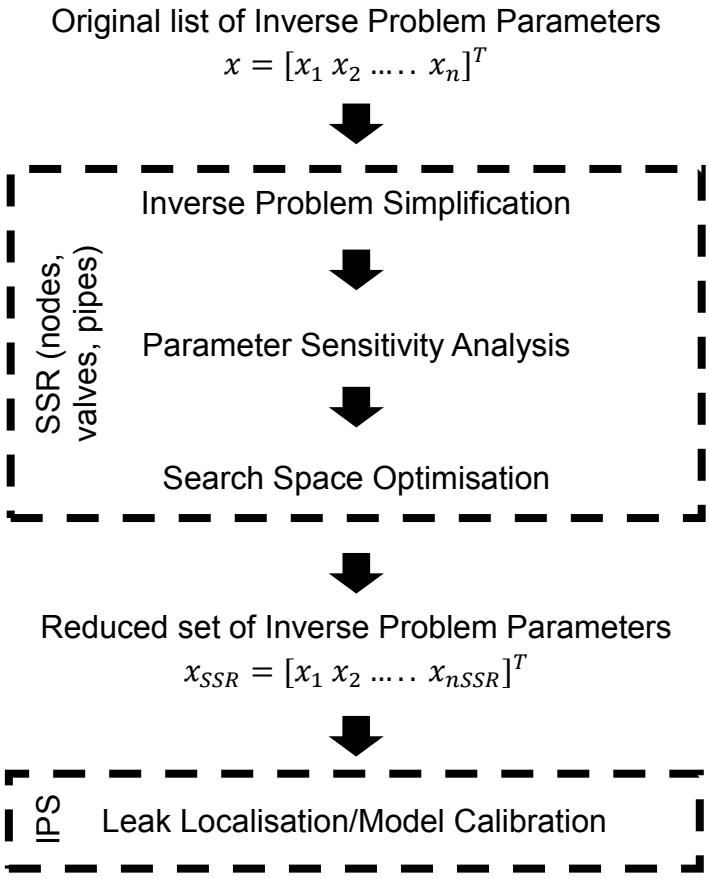

Figure 3.1. Overview of the decision-support framework for solving *inverse problems* in WDN modelling based on search space reduction.

The developed search space reduction technique allows the infeasible solutions of the randomly generated initial GA population, to move towards the feasible region, which contracts the search space. Consequently, in a reduced search space, the optimisation analysis for solving the *inverse problem* starts with a better population of solutions and is allowed to converge faster and more effectively towards the global optimum. In addition, the unnecessary simulation of solutions that cause no impact on model fitness is avoided. Three types of decision variables are considered, which are associated with: (i) Leak Detection and Localisation, and (ii) Model Calibration. These include:

(a) The emitter coefficient values of nodes in the WDN hydraulic model used for leak detection and localisation;

(b) The status of valve components used in model calibration for detecting faults in the hydraulic model or the real system, associated with unknown closed/open throttle valves;

(c) The roughness coefficient value of pipe components used in model calibration for determining pipe friction factors and, therefore, estimating velocities of flow paths in the real system.

For each decision variable type the SSR stage is performed in three steps:

(1) Inverse Problem Simplification (IPSI);

(2) Parameter Sensitivity Analysis (PSA); and

(3) Search Space Optimisation (SSO).

The IPSI in Step 1, takes into account prior information, expert knowledge of the system and some assumptions, in order to reduce the *inverse problem* size. During PSA in Step 2, a number of scenarios are simulated. Considering the configuration and characteristics of pressure sensors in the WDN, any parameters or parameter values that cause no impact on the WDN model hydraulics are falsified. Finally, in SSO in Step 3, another set of scenarios is analysed and compared to the available observations. Those that do not affect the WDN model fitness are eliminated. Ultimately, the SSR stage ensures that the size of the original *inverse problem* is minimized, whereby the list of decision parameters for each parameter type is restricted to a reduced:

(a) Number of candidate leak locations and range of possible flow values,

(b) Number of candidate throttle valve locations with uncertain status,

(c) Number of candidate pipes and range of possible roughness values.

3.3 Inverse Problem Simplification

3.3.1 Introduction

In Step 1 of the SSR stage the number of decision variables of the original *inverse problem* is minimized. This is achieved through a simplification process that considers prior information and knowledge of the system, along with assumptions that either: (a) exclude decision variables of the problem, or (b) group them based on their characteristics. This results in a simplified set of parameters, which leads to a reduction in the problem dimensionality.

3.3.2 Leakage Nodes

A node is normally considered a potential leak location if there is no available data associated with its actual demand value. Therefore, all nodes in a WDN

model (linked to pipes, valves, pumps and tanks) could be potential leak locations (i.e., decision variables) resulting in a vast search space. In reality the majority of leaks happen on pipes, but the model normally assigns aggregated demands to nearby nodes, hence it is assumed here that leaks are located on network nodes only. In a WDN model not all nodes represent areas of aggregated demand. For example, in EPANET modelling software two nodes are added to a pipe section just upstream and downstream of a valve to indicate its boundaries, without assigning any demand to them. A similar process is used to indicate the physical limits of a pumping station. Although it is true that valves can leak due to a weak stem, the losses are normally insignificant relative to the inlet flow, or to the undetected leaks. In Step 1 of the SSR stage the number of potential leak locations is reduced, assuming that leaks only happen on pipes. Thus, in IPSI all nodes associated with non-pipe components are excluded (Figure 3.2).

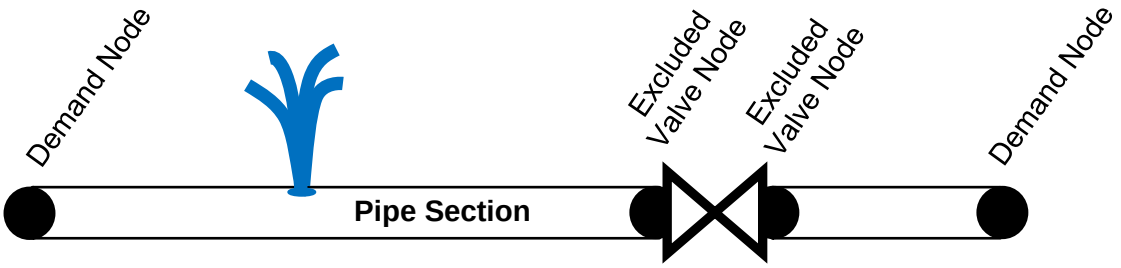

Figure 3.2. Reduction via Inverse Problem Simplification for the candidate leak locations.

3.3.3 Valve Components

A throttle valve is considered a calibration parameter (i.e., a decision variable) during model calibration if there is an associated uncertainty in its status, i.e., either open or closed. It is also possible that a throttle valve can be partially closed/open, however, this issue is beyond the scope of this thesis, due to the insignificant hydraulic impact of such condition, as opposed to the case when valves are fully closed/open. Nevertheless, most utilities have no live records for the status of all valves in the WDN. Consequently, all valves in a WDN model could be locations of uncertain status. This increases the complexity of the *inverse problem*. On the other hand, it is possible to falsify candidates and simplify the problem dimensionality, based on their location in the system.

In Step 1 the number of candidate valves with uncertain status is minimized by considering the available information associated with the valve location, as well as expert knowledge of the system. Any valve where live or recent data exists for its status can be excluded from the search space. The same stands for any valve that is located on a flow monitored pipe as its status can be determined indirectly, from the flow measurement value. If the flow is zero, then, the valve is closed, otherwise it is open. Furthermore, the status of valves that define the limits of a DMA is known, unless any unknown open boundary valves exist. Except from that, it is unlikely that a valve is unknowingly closed if it is located on a dead end pipe, or branched part of the WDN (Figure 3.3), unless the properties that are supplied from that pipe are fed from an unknown source, which is therefore not presented in the hydraulic model, or the Geographic Information Systems map. In a real system, unknown closed valves on any branched pipe would be sensed by the customers. This is because in such a situation the WDN part/segment would be isolated, causing an interruption to supply. Therefore, it is expected that any customer experiencing this situation would report it to the water utility.

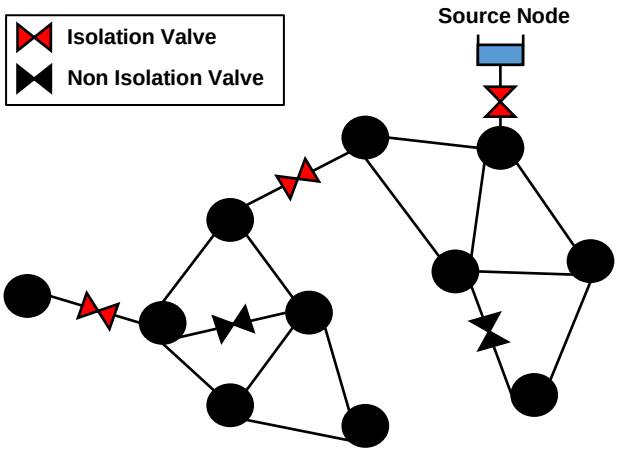

Figure 3.3. Network representation of valves that lead to isolation.

In IPSI, the number of candidate valves with unknown status can be initially minimized, assuming that isolation valves, i.e., those located on branched parts of the WDN, or dead end pipes, are open. Thus, only valve components that are located on loops are included in the search space. In addition, from those that are located on loops some can be eliminated based on the consequences that a

change in their status would cause to the system's functionality. For instance, the status PRVs, or valves that are located on by-passes can be determined indirectly from the pressure variability in the WDN. During normal conditions active PRVs are always open, whereas inactive PRVs are closed only if an open by-pass exists. Valves on by-passes are normally closed, as otherwise the pressure would not be managed by the PRV. In general, any valve whose status is uncertain is included in the search space for model calibration if:

(a) There is no live/recent data for its status.
(b) It is not located on a pipe with a flow meter.
(c) It does not lead to boundary changes between two or more DMAs.
(d) It does not lead to isolation of a WDN part/segment.
(e) It is not an active PRV.
(f) It is not located on a by-pass of an active PRV.

3.3.4 Pipe Roughness Groups

During the model calibration process any pipe in the WDN model can be calibration parameter (i.e., a decision variable) if there is no available information about its roughness coefficient values. The only way to know (or verify after model calibration) the true roughness value is to dig the ground and record data that can be used to determine it. However, in reality it is impossible to physically measure every pipe in the WDN as this would lead to huge costs for network operators. The pipe roughness coefficients can be estimated by solving the *inverse problem*, based on a number of collected measurements. On the other hand, with so many decision variables the search space of the optimisation process becomes huge.

The number of decision variables can be reduced by considering similarities in pipe characteristics, based on the available information, and dividing them into different groups. The grouping process in Step 1 considers three main characteristics of pipes, including: (a) Material; (b) Age; and (c) Diameter. During IPSI, the WDN model pipes are firstly divided into, PM, classes based on their material as pipes of same material share the same range of possible roughness values. If age information is available it is possible to sub-divide each material class, pm, into, PA_{pm}, sub-classes. This is because pipes of same material and age are assumed to share a similar rate of change in their roughness coefficients.

However, not all water companies hold such records for the state of pipes in the WDN. Thus, this step is optional as it can only be carried out if this information is available, otherwise all pipes are assumed to have a similar age, i.e., $pa_{pm} = 1$. Then, it is possible to separate the resulting sub-classes into $PG_{pa_{pm}}$ groups based on the diameter. Pipes of same diameter are expected to affect head loss in a similar way due to their hydraulic boundaries on flow. Ultimately, the list of decision variables, DV, for the calibration process is reduced to:

$$DV = \sum_{pm=1}^{PM} \sum_{pa=1}^{PA} PG_{pa_{pm}} \quad (3.1)$$

3.4 Parameter Sensitivity Analysis

3.4.1 Introduction

Following Step 1, a reduced set of decision variables is established, which is used as an input to Step 2 of the SSR stage. In this step, a sensitivity analysis is carried out for each parameter (and for each parameter type). This is done to assess their effect on the simulated pressure, as pressure monitoring devices are used in a larger extent for data collection rather than flow metering devices. The aim is to identify the parameters and parameter values that cause little or no impact on the simulated pressure, relative to the defined boundary conditions. They are then removed from the search space. As a consequence, this will lead to avoidance of unnecessary simulation of solutions that do not affect the model fitness, i.e., the objective function, when solving the *inverse problem*. It also allows to reduce the *inverse problem* size and help with the non-uniqueness issue.

During the PSA in Step 2, the accuracy range of the installed in pressure sensors the WDN is initially taken into account to set a threshold value for a minimum detectable pressure response. Each time, the value of a parameter, is systematically adjusted and the resulting pressure response is compared against the pressure from the defined boundary conditions. The philosophy of this step in SSR stage is that if the pressure perturbation from a simulated event is similar or less than the threshold detectable pressure response set from the accuracy range of pressure devices, then, the event will remain undetectable regardless of the distance from the sensor. This is because the uncertainty in measurement is

similar or larger than the impact caused by the change in the WDN model. Based on the decision variable type, the PSA ultimately determines:

(a) The Minimum Detectable Nodal Leakage (MDNL)

(b) The Detectable Valve Locations (DVLs)

(c) The Detectable Pipe Roughness Coefficients (DPRCs)

The resulting sets of decision variables and the corresponding range of values are subsequently considered as an input for Step 3 of SSR stage.

3.4.2 Minimum Detectable Nodal Leakage

Depending on the sensor configuration there is limited observable WDN space, i.e., the length of pipes that can be monitored for leakage. If the leak distance from the sensor increases, the minimum detectable flow increases due to the lower head loss. Furthermore, all devices are accurate within a specified range indicated by the manufacturers. Thus, if the pressure perturbation from a leak is below the device's accuracy range, the event will remain undetectable regardless of the distance from the sensor. Based on the number, location and accuracy range of sensors there is a minimum detectable flow for each location, which establishes a lower bound when detecting leaks.

The Minimum Detectable Nodal Leakage (MNDL) process starts by imposing the boundary conditions for the leak-free scenario and analysing the resulting pressure ($\widehat{p_{0s}}$) at nodes, s, where sensors are present. Then, a leak with large flow, Q (specified by the user), relative to the system's average demand over 24 hours is simulated at every potential leak location, l, and the pressure response, $p_{Q,l,s}$, at location s is recorded. The number of potential leaks is equal to the number of nodes remaining after implementing Step 1 (Section 3.3.2). Leaks are simulated as emitters following Equation 2.12 (Chapter 2). The pressure residuals between a no-leak and leak scenario are determined for all possible leak nodes across all the simulation time steps. These differences between the boundary conditions and leak response pressure at sensor nodes are then averaged over all time steps, using:

$$r_{Q,l}(t) = \begin{bmatrix} p_{Q,l,1}(t) - \widehat{p_{01}}(t) \\ \vdots \\ p_{Q,l,s}(t) - \widehat{p_{0s}}(t) \end{bmatrix} \tag{3.2}$$

Each time, the simulated leak flow, Q, is systematically reduced by adjusting the value of c until the residuals across all sensors do not exceed their accuracy range, ε, (e.g., typically ±0.1% measurement error of full scale: 0-10 bar). The value of ε sets the threshold value of minimum detectable pressure response. Ultimately, this establishes the MDNL, $Qmin_l$, for each potential leak node in the WDN model.

$$MDNL = \begin{bmatrix} Qmin_1 \\ \vdots \\ Qmin_l \end{bmatrix} \tag{3.3}$$

3.4.3 Detectable Valve Locations

Similar to the case of MDNL, the sensor number, location and accuracy determines the observable WDN space where changes in the status of valves can be identified. Typically, the pressure impact from a change in valve status is local. Furthermore, in current WDNs, the main source of information in identifying such changes comes from pressure measurements, due to the lack of flow measurements. Based on the same philosophy as in the case of a leak, if the pressure perturbation from a change in valve status is below the device's accuracy range, the event will remain undetectable.

The analysis of Detectable Valve Locations (DVLs) starts by analysing the boundary pressures ($\widehat{p_{0s}}$) at nodes, s, resulting from the situation when no changes in the status of any valve are made. Then, each time, the status of candidate valve, k, is changed, and the pressure response, $p_{k,s}$, is recorded. If a valve is originally set as closed in the WDN hydraulic model, then, its status is changed to open, or vice versa. The number of candidate valves is equal to the remaining valve components after implementing IPSI in Step 1 (Section 3.3.3). The pressure residuals between a no-change case and change in valve status are determined for all candidate valves across all the simulation time steps as the difference between the boundary conditions and the valve status change response pressure at sensor nodes, respectively, using:

$$r_k(t) = \begin{bmatrix} p_{k,1}(t) - \widehat{p_{01}}(t) \\ \vdots \\ p_{k,s}(t) - \widehat{p_{0s}}(t) \end{bmatrix} \tag{3.4}$$

Any valve that produces a response in simulated pressure across all sensors, of less than, or equal to the accuracy range, ε, is classified as unobservable and is removed from the search space. This is because detection of a status change or error cannot be actually performed, as the effect on simulated pressures is less or similar to the measurement uncertainty. At the end of this procedure, the list of all DVLs in the WDN model is established, which is considered in the next step of the SSR.

3.4.4 Detectable Pipe Roughness Coefficients

Water utilities hold records for the range of possible roughness coefficient values for each pipe material class. This defines the range of possible discrete values for each pipe during model calibration, based on its material. However, changing the roughness coefficient value, λ, of a single pipe, or new pipes in the WDN model would normally cause little or no change in the simulated hydraulics. Therefore, typically, during model calibration changes are performed in number or groups of pipes each time, where an observable impact on the simulated hydraulics is produced. Before outlining the approach used for pipe components it is important to clarify that the decision variable λ used here, for specifying the pipe condition is the value used to determine the friction factor and not the friction factor per se. The number of possible roughness values, Λ, for each group, pg, depends on the material class, pm, where those of same material class share the same range of possible roughness values, RV_{pm}. The range of possible roughness values for each material class is established from:

$$RV_{pm} = \begin{bmatrix} \lambda_{1,1} & \cdots & \lambda_{\Lambda,1} \\ \vdots & \ddots & \vdots \\ \lambda_{1,PM} & \cdots & \lambda_{\Lambda,PM} \end{bmatrix} \tag{3.5}$$

Similarly to the case of nodes and valves, there is a limited number of pipes where calibration can be performed, depending on the sensor configuration. A pressure perturbation as a result of an adjustment in the pipe group's roughness value of less than, or equal to the device's accuracy range, ε, cannot affect the fitness of the hydraulic model. The Detectable Pipe Roughness Coefficients (DPRCs) analysis applies the same philosophy as in sections 3.4.2 and 3.4.3. The process starts by setting the roughness value, $\lambda_{pg,pm}$, of all pipes in group, pg, of material class, pm, to a value equal to the minimum nominal value, λmin_{pm} and analysing

the resulting pressure ($\widehat{p_{0s}}$) at sensor nodes, s. This sets the boundary conditions for the original case, i.e., when the pipes were new. Moreover, it establishes the minimum possible value, $\lambda_{1,pg,pm}$, for the pipe group. The number of pipe groups for each pipe material is equal to the $PG_{pa_{pm}}$ groups, following IPSI in Step 1 (Section 3.3.4). The DPRC analysis compares the simulated pressures from boundary conditions with those resulting after the adjustment of the roughness values for the pipe group. To determine the DPRCs of any pipe group, pg, of material, pm, each time the roughness coefficient value, $\lambda_{pg,pm}$ of all pipes that belong in that group is changed to test all possible values in RV_{pm} and the pressure response, $p_{\lambda_{pg,pm},s}$, is recorded. The pressure residuals between the minimum nominal roughness value and adjusted case are determined for each group across all the simulation time steps as in the previous sections using:

$$r_{\lambda_{pg,pm}}(t) = \begin{bmatrix} p_{\lambda_{1,pg,pm},1}(t) - \widehat{p_{01}}(t) \\ \vdots \\ p_{\lambda_{\Lambda,pg,pm},s}(t) - \widehat{p_{0s}}(t) \end{bmatrix} \tag{3.6}$$

Any $\lambda_{pg,pm}$ that produces a simulated pressure response across all sensors, of less than or equal to the accuracy range, ε, is removed from the list of possible values for each pipe group. If none of the tested values in RV_{pm} causes a residual, $r_{\lambda_{pg,pm}}$, larger than ε, the whole pipe group is removed from the search space. Following DPRC analysis any unobservable pipe groups are excluded, as their inclusion in the search space will lead to simulations that cause no change in the model fitness during model calibration. Ultimately, this leads to a new reduced list of decision variables, $DV = PG\ x\ PM$, with their corresponding range of possible roughness coefficient values, $RV_{pg,pm}$. In addition, for each observable pipe group the range of possible roughness values, $DPRC_{pg,pm}$, is established as:

$$DPRC_{pg,pm} = \begin{bmatrix} RV_{1,1} & \cdots & RV_{1,PM} \\ \vdots & \ddots & \vdots \\ RV_{PG,1} & \cdots & RV_{PG,PM} \end{bmatrix} \tag{3.7}$$

This set of decision parameters is subsequently considered in Step 3 of SSR.

In Step 3, which is explained in the next section, final search domain for all decision variable types is established corresponding to the *inverse problem* being solved.

3.5 Search Space Optimisation

3.5.1 Introduction

In Step 3 of the SSR a number of optimisation problems are formulated and solved with the ultimate purpose to minimize the number of decision variables and the range of possible values. A Genetic Algorithm (GA), that is a computationally fast and elitist evolutionary algorithm, is used (Deb et al., 2002). Both flow and head observations from pressure measurements are used along with the remaining parameters from PSA in Step 2. The output from the SSO process is the reduced final list of decision variables, x_{SSR} (Figure 3.1), before solving the *inverse problem*. The problem is formulated using the two different sets of constraints, described in Section 2.2.2 (Chapter 2).

Although optimisation is formulated differently depending on the decision variable type, in all cases this is done with a fixed number of decision variables. This is explained in more detail in subsections 3.5.2 – 3.5.4. Based on the decision variable type, the SSO process determines:

(a) The total water losses and the maximum number of possible leaks in the WDN.
(b) The maximum number of possible closed valves in the WDN.
(c) The number of pipe groups that can be calibrated.

3.5.2 Leak Locations and Range of Flow Values

3.5.2.1 Overview

The SSO in Step 3 of the SSR is divided into two parts, i.e., Part I and Part II. Here, different optimisation problems are solved to minimize the number of candidate leak locations and the range of possible flow values. For any possible leak two decision variables are defined, one for the discrete location and one for the continuous flow value (Wu et al., 2010).

In Part I optimisation analysis the decision variables include the remaining nodes from PSA in Step 2, optimised as a single emitter coefficient. Using only observed flow measurements, the aim is to detect the total water losses in the WDN based on the simplest case scenario when the entire water loss volume is associated

with a single leak. The outcome of Part I defines the maximum emitter flow for each potential leak location, which reduces the range of flow values for the subsequent analyses (Figure 3.4). Moreover, it allows to falsify any candidate nodes with higher MDNL values than the losses detected in Part I.

In Part II a second optimisation problem is solved with an updated list of candidates and range of flows. Here, the aim is to estimate the maximum number of possible leak locations occurring in the WDN, i.e., the most complex case scenario. Both flow and head (calculated from pressure and elevation) observations are considered during the optimisation analyses. Using the value of the detected total water losses from Part I, and the identified maximum possible leak scenario, the outcome defines the minimum emitter flow for each remaining candidate. Consequently, it leads to further reduction in the number of possible flow values for each potential leak location.

Based on the output of Part II optimisation analyses, the final search domain is defined. It is comprised of the reduced list of potential leak locations and the corresponding reduced range of flow values (including a zero leakage value). The optimisation problems are solved similarly in both parts. The optimisation in each part (Figure 3.4) is formulated as follows:

Search for: $\vec{X} = (LN_i^n, c_i^n);$ $LN_i^n \in J;$ $c_i^n \in K;$
$n = 1, \dots, NLeak;$ $i = 1, \dots, NIndex$ (3.8)

Minimize: $F(\vec{X})$ (3.9)

Subject to: $[0, \ \underline{c^n} \leq c_i^n \leq \overline{c^n}]$ (3.10) $P > 0$ (3.11) $NLdup^n = 0$ (3.12)

Where: LN_i^n is the index for node i, representing the location of the possible leak n;
c_i^n is the emitter coefficient (equivalent to c_i in Equation 2.12 – Chapter 2), for node i, representing the size of the possible leak n;
J is the set of potential leak locations for any possible leak size;
K is the range of possible values for any potential leak location;
$NLeak$ is the number of possible leak locations to be identified;

$NIndex$ is the number of the candidate nodes for any possible leak size;

$\underline{c_i^n}$ and $\overline{c^n}$ are the minimum and maximum emitter coefficients, respectively, for the node i for the possible leak n along with a possible value of 0 that can be selected to indicate a no-leak case;

P is the pressure at any WDN node;

$NLdup^n$ is the number of the duplicate nodes that are identified as leakage emitters in the same solution.

The constraints given by Equations 3.11 and 3.12 are handled by introducing an artificial penalty for violating the constraint. They penalize solutions by reducing their fitness values in proportion to the degrees of constraint violation. These are necessary for the GA to search for good and realistic solutions. Thus, any solutions that may cause negative pressures in the system or those where a node is identified as a location of multiple leaks if $n > 1$, are avoided.

3.5.2.2 Part I

In Part I a single leak scenario ($n = 1$) is assumed, i.e., c_i^1, to only detect the total water losses in the WDN. Therefore, no information is sought about the precise leak location(s). A series of short (user specified) optimisation analyses, Z, are undertaken, where at each analysis, z, the emitter coefficient, c_i^1, is allowed to vary within a specified range d_z (user specified), corresponding to the maximum and minimum flow bounds, while zero is not included as an option. The first analysis, $z = 1$, begins with, c_i^1, equal to the maximum difference between the observed and simulated system's demand throughout the 24 hours. This includes both the error caused by leakage as well as the modelling and measurement uncertainty. Each solution's c_i^1 values are allowed to vary within a wide range, e.g., $d_1 = \pm 50\%$ relative to the c_i^1 value that initiates the process. The wider range of values at the start allows to compensate for premature convergence caused by large differences in pressures across the WDN that may result in different optimal c_i^n values for the same leak flow. Furthermore, to mitigate the impact caused by the idealized pressure exponent value of a (Equation 2.12 - Section 2.5.4.2.1, Chapter 2) that is considered here as part of the leak orifice equation.

When a fit solution is found, the emitter value for the next analysis is updated with the optimal solution found. In addition, the range of adjustment, d_{z+1}, is reduced to fine tune the solution for detecting the total water loss. For example, if the optimal solution in analysis, z, is fitter than the previous, $z-1$, the identified emitter is used to specify c_i^1 for the next one, $z+1$. For subsequent analyses the bounds of the c_i^1 initiating the optimisation process are systematically reduced to eventually reach, $d_Z = \pm 1\%$ (e.g., ±25%, ±10%, ±5% and ±1%). Those values are relative to the optimum solution of the previous analysis, which allows for both global and local exploration of the search space. The optimisation minimizes the weighted sum of squared flow differences, given by:

Minimize: $$F(\vec{X}) = \sum_{t=1}^{T}[\sum_{f=1}^{F} \frac{w_f\left(\frac{Qo_f(t)-Qs_f(t)}{Qpnt}\right)^2}{F}] \quad (3.13)$$

Where: $Qo_f(t)$ is the observed flow (l/s), in link f at time t;

$Qs_f(t)$ is the simulated flow (l/s), in link f at time t;

$Qpnt$ is the measurement error (l/s), which converts flow differences into a dimensionless value based on the meter accuracy. This is done so that the flow error can be compared to the pressure error, which is used in the next part when pressure measurements are considered, given by:

$$Qpnt = \bar{Q}_{in} Accuracy_Q \quad (3.14)$$

$\bar{Q}_{in}$ is the average global system demand;

$Accuracy_Q$ is the flow reading percentage error as indicated by the meter manufacturer;

F is the number of metered pipes, used to mitigate the impact caused by great differences in the number of available pressure and flow measurements;

w_f is a weighting factor for observed flows, which is used to differentiate the value of information gained by the measurements during different times of the day, such as the hours of low demand.

Following all optimisation analyses the optimal c_i^1 corresponds to the leak flow value (l/s), which represents the global water losses without considering the local

modelled pressure variation, which still does not match the observations. The leak flow value is used in Equation 2.12 (Chapter 2) to calculate the final emitter coefficient based on the average pressure obtained from all sensor measurements and assuming an average value of $a = 0.5$ across the system. As it is based on a single orifice, it establishes an upper bound flow, $\overline{c^n}$, for all potential leaks. Any potential leak node with a larger MDNL than the leak flow, $Q_{\overline{c^n}}$, as a result of $\overline{c^n}$ is removed from the search space. Therefore the potential leak locations are redefined, producing the final list of candidate nodes.

3.5.2.3 Part II

In Part II a series of n leak scenarios are simulated to estimate the maximum number of possible leaks occurring in the WDN and ultimately define the minimum emitter flows for the remaining leak locations. The total number of leak scenarios, $NLeak$, is derived from $Q_{\overline{c^n}}$ and the average MDNL from the remaining candidates, using:

$$NLeak = \frac{Q_{\overline{c^n}}}{MDNL\ avg.} \tag{3.15}$$

Optimisation analyses are carried out to minimize the weighted sum of squared differences between observed and simulated values for both heads and flows as follows:

Minimize: $$F(\vec{X}) = \sum_{t=1}^{T}[\sum_{s=1}^{S} \frac{w_s\left(\frac{Ho_s(t)-Hs_s(t)}{Hpnt}\right)^2}{S} + \sum_{f=1}^{F} \frac{w_f\left(\frac{Qo_f(t)-Qs_f(t)}{Qpnt}\right)^2}{F}] \tag{3.16}$$

Where: $Ho_s(t)$ is the observed head (m) of node s at time t;

$Hs_s(t)$ is the simulated head (m) of node s at time t;

$Hpnt$ is the measurement error (m), which converts head differences into a dimensionless value based on the sensor's reading accuracy, given by:

$$Hpnt = \overline{P_s} Accuracy_p \tag{3.17}$$

$\overline{P_s}$ is the average observed pressure (m) at sensor nodes;

$Accuracy_p$ is the sensor reading percentage error as indicated by the manufacturer;

S is the number of sensors, used similarly as F;

w_s is a weighting factor for observed heads, used similarly as w_f; All flow-related symbols are similar as in Part I formulation.

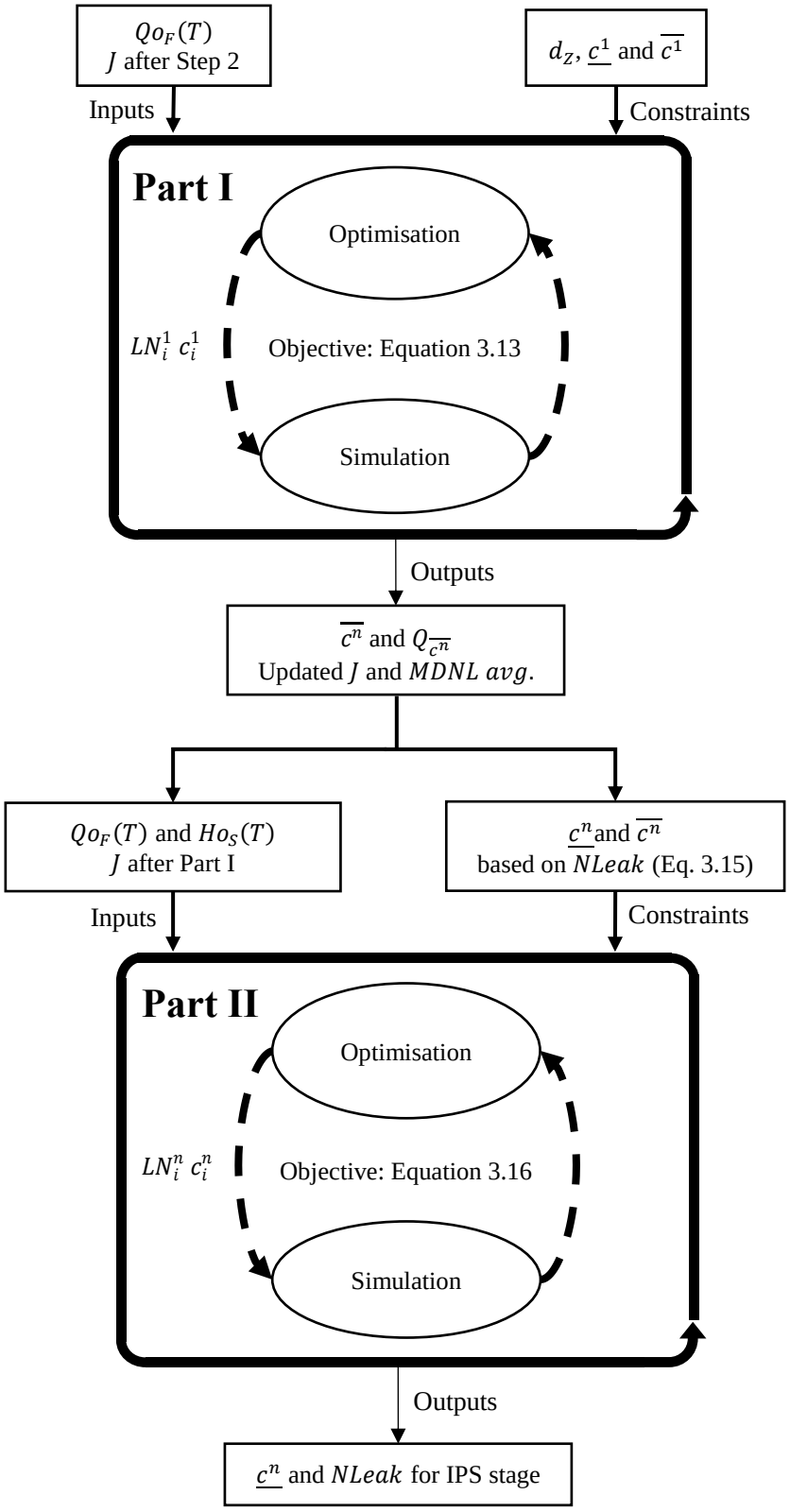

Figure 3.4. Search Space Optimisation for Leakage Nodes.

The fittest n scenario establishes the minimum emitter, $\underline{c^n}$, for the candidate leak locations as the total leakage from the identified leaks in the optimal solution should equal the total water losses, $\overline{c^n}$, determined in Part I. The range of flows for any node with larger MDNL than $\underline{c^n}$ is adjusted within its MDNL and $\overline{c^n}$. The process flow chart for Step 3 is presented in Figure 3.4.

3.5.3 Closed Valves

A series of scenarios, v, are simulated with the ultimate purpose to estimate the maximum number for possible closed valves in the WDN. The outcome of this process also corresponds to the number of possible open valves, thus, allowing to ultimately detect both unknown closed and open valves in the WDN model. The total scenarios $NValve$ is equal to the number of candidate valves remaining from PSA in Step 2. The optimisation framework with fixed number of decision variables, associates one parameter for any possible closed valve that also corresponds to the scenario, v. Using only observed heads from pressure measurements, the aim is to estimate the maximum number of possible closed valves in the WDN, which limits the search to a defined space. The decision variables of the optimisation problem include the set of remaining valve locations, V, following Step 2, which is formulated as:

Search for: $\vec{X} = (CV_k^v); \quad CV_k^v \in V;\ v = 1, \dots, NValve;\ k = 0, \dots, VIndex$ (3.18)

Minimize: $F(\vec{X})$ (3.19)

Subject to: $P > 0$ (3.20) $NLdup^v = 0$ (3.21)

Where: CV_k^v is the index of locations k corresponding to the possible closed valve v;

V is the set of candidate locations for any possible closed valve;

$NValve$ is the number of possible closed valves to be identified;

$VIndex$ is the number of the candidate locations for any possible closed valve;

P is the head at any node;

$NLdup^v$ is the number of the duplicate valve locations that are identified as closed in one solution for the possible valves v. The number 0 is also included in the range of index values, k, for a

closed valve location. This is added for more flexibility so that a solution with less than v closed valves is also possible. The constraints given by Equations 3.20 and 3.21 are equivalent to Eq. 3.11 and 3.12 for the GA to search for good solutions.

The optimisation analyses are carried out to minimize the weighted sum of squared differences between observed and simulated head values, given by:

Minimize: $$F(\vec{X}) = \sum_{t=1}^{T}[\sum_{s=1}^{S}\frac{w_s\left(\frac{Ho_S(t)-Hs_S(t)}{Hpnt}\right)^2}{S}] \quad (3.22)$$

The abbreviations have a similar meaning as in Section 3.5.2. The fittest scenario establishes the maximum number of possible closed valves, $NValve$, in the WDN. This is used during the IPS stage and allows to identify both unknown closed and open valves. The process flow chart for the valve search space optimisation is presented in Figure 3.5.

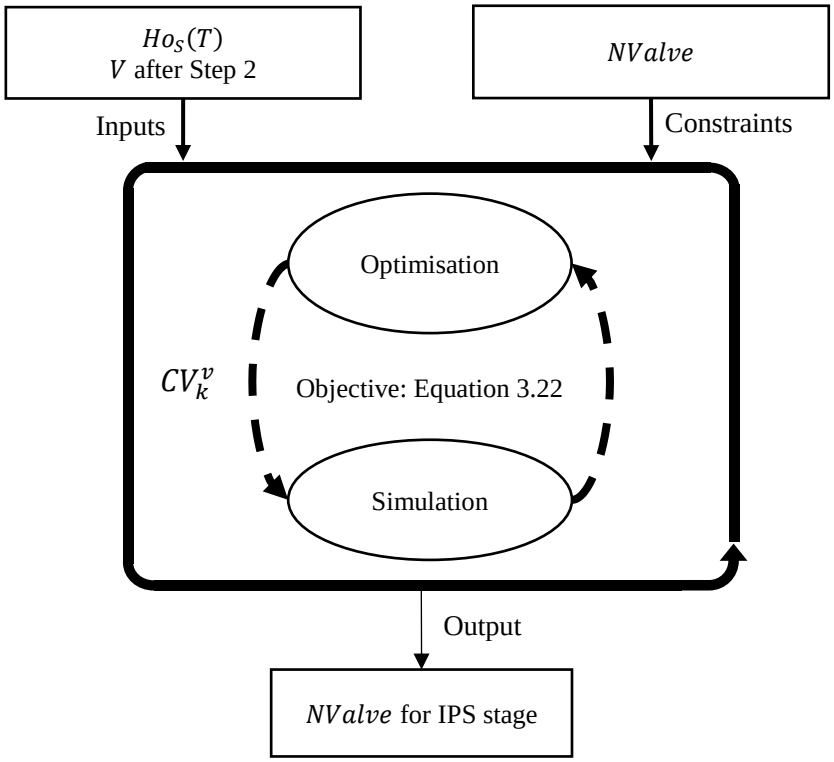

Figure 3.5. Search Space Optimisation for Valve Components.

3.5.4 Pipe Roughness Calibration Groups

Step 3 for Pipe Roughness Calibration Groups is also associated with a series of scenarios, $NGroups$. A number of short (user specified) optimisation analyses, Z, are simulated at each scenario, g, with the ultimate purpose to estimate the maximum number of candidate pipe roughness calibration groups (Figure 3.6). This limits the calibration procedure to a defined number of pipes and possible roughness values. The optimisation framework with fixed number of decision variables, associates one decision variable for the roughness value of any pipe group that also corresponds to the scenario, g. This represents a group of discrete pipes whose discretized roughness value is adjusted. The decision variables include the remaining pipe groups and the corresponding roughness coefficients from Step 2 (section 3.4.3). Using only observed heads from pressure measurements (as in 3.5.3), the aim during each scenario is to two-fold:

(a) To evaluate the improvement caused on the model fitness as a result of any changes in the roughness value of each pipe group.

(b) To rank each considered pipe group according to its sensitivity on the objective function value.

This allows to falsify the groups of pipes that cause no impact on the model fitness, and eventually reduce the problem dimensionality. At the end of Step 3 the final search domain for the calibration of pipe roughness is defined, comprised of the reduced list of pipe groups and their corresponding range of possible roughness values. The optimisation problems are formulated as:

Search for: $\vec{X} = \left(\lambda_j^g\right);$ $\lambda_j^g \in RV_{pg,pm};$

$g = 1, \dots, NGroups;$ $j = 1, \dots, RIndex;$ (3.23)

Minimize: $F\left(\vec{X}\right)$ (3.24)

Subject to: $\left[\underline{\lambda_j^g} \leq \lambda_j^g \leq \overline{\lambda_j^g}\right]$ (3.25) $P > 0$ (3.26)

Where: λ_j^g is the roughness coefficient value for the group of pipes g corresponding to index j in set $RV_{pg,pm}$ associated with pg of material class pm;

$RV_{pg,pm}$ is the set of possible roughness values that belongs in $DPRC_{pg,pm}$ for the group g associated with pg and material class pm;

$NGroups$ is the total number of pipe groups to be calibrated;
$RIndex$ is the number of possible roughness values for pipe group, g, associated with pg and material class, pm;
P is the head at any node.

The process starts by imposing a boundary objective function value, OF_0, associated with the no-change scenario, i.e., if no changes are made in the model. The first scenario, $g = 1$, considers a single pipe group and $Z = NGroups$ optimisation analyses are carried out. At each analysis, z, a different group of pipes, pg, is considered for optimisation of its roughness value and the resulting change in objective function, OF_z, is analysed. This is done to assess the improvement on the model fitness, i.e. the reduction in the objective function value, resulting from changes in the roughness coefficient value of a single pipe group. The process is undertaken for each pipe roughness group separately. The fitness improvement, fi_z, for each z, which also represents a measure of sensitivity with respect to the objective function, is given by:

$$fi_z = OF_0 - OF_z \tag{3.27}$$

The optimisation minimizes the weighted sum of squared head differences, given by Equation 3.22 in section 3.5.3. Following all optimisation analyses the set of fitness improvement values FI^g for all tested pipe groups in scenario, g, is established, given by:

$$FI^g = \begin{bmatrix} fi_1 \\ \vdots \\ fi_Z \end{bmatrix} \tag{3.28}$$

The list FI^g ranks each pipe group considered during the optimisation analyses in a descending order, according to its sensitivity on the objective function. The pipe group that causes the largest fitness improvement (i.e. the most sensitive pipe group) is added as a decision variable in the next scenario, $g = g + 1$. It also populates the list GS, which represents the decision variables ordered by their sensitivity on the OF. Subsequently, the remaining groups are optimised each time separately, in combination with the chosen pipe group in $Z = NGroups - g$ optimisation analyses. The same philosophy is applied for the rest of the procedure until all scenarios $g = NGroups$ are tested and GS is fully populated. At each subsequent scenario the number of decision variables, g, is increased

by one, by adding the most sensitive group from the ranked FI^g and the resulting fitness improvement is analysed for all remaining pipe groups. This is done to test the fitness improvement (Eq. 3.27) with an extended set of decision variables where the optimised pipe roughness calibration groups are combined differently. Conversely, the number of analyses, Z, at each scenario is decreased by one, as the chosen groups cannot be duplicated during the optimisation. When all scenarios, $g = NGroups$ are completed the scenario with the largest fitness improvement overall, establishes the chosen pipe groups from GS that are considered for calibration, while the rest are falsified, as a result of no, or worse fitness change. Ultimately this leads to a reduced list of highly sensitive pipe groups and their corresponding roughness coefficient values that serve as an input for the IPS stage. The process flow chart for the pipe group search space optimisation is presented in Figure 3.6.

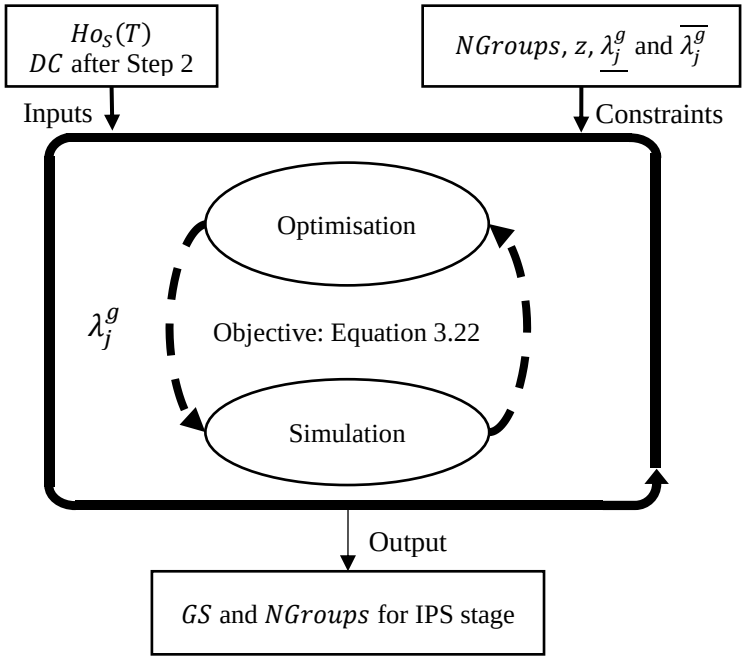

Figure 3.6. Search Space Optimisation for Pipe Groups.

3.6 Inverse Problem Solving

At the IPS stage, the ultimate *inverse problem* is formulated. Depending on its aim, i.e., leak detection and localisation or model calibration, the optimisation is formulated with the appropriate decision variables. This is explained in more detail in the next chapter, where methods for both leak detection and localisation

as well as model calibration are presented. The decision variables include the list of parameters identified in SSO of the SSR stage. During the IPS stage both flow and head observations are used to minimize Equation 3.16.

In this research during the hydraulic model calibration the optimisation determines the roughness coefficient value of each pipe group and/or the number and locations of closed valves (see Section 4.5.5). Therefore the problems formulated in sections 3.5.3 and 3.5.4 can be either solved separately, or together. For leak detection and localisation, here, the optimisation detects and localises the emitter flow values in the WDN (see Section 4.4.4). If necessary this stage in leak detection and localisation can be combined with hydraulic model calibration to identify any existing errors carried from previous model building and calibration procedures.

3.7 Summary and Conclusions

In this chapter, a search space reduction technique was presented to overcome the ill-posedness of *inverse problems* that concern, applications for either model calibration or leak detection and localisation. The technique is part of the proposed two-stage decision-support framework for *solving inverse problems* in WDN modelling and is implemented in three steps: (1) Inverse Problem Simplification, (2) Parameter Sensitivity Analysis and (3) Search Space Optimisation. In Step 1 prior information and knowledge of the system is used to simplify the problem and eliminate known, insensitive and unnecessary decision variables. In Step 2 the remaining decision variables are analysed with respect to their impact on the WDN model hydraulics and compared against the realistic boundary conditions. Any insensitive parameters are removed from the search space. Finally, in Step 3 the search space is further restricted based on additional optimisation analyses. This step establishes the final list of decision parameters prior to solving the *inverse problem*.

The main conclusions from this chapter are as follows:

- Solving an *inverse problem* in WDN modelling is usually a complex task. Any available information that is directly or indirectly related to its aim should be used, even if it does not seem important initially.

- The technique can systematically reduce the number of decision variables and the range of possible values, considering the error in the available data. This is the key to reducing the *inverse problem* search space, as it provides a threshold value for the observable parameters.
- Several improvements can be made to condition an ill-posed problem, e.g., reformulation of the problem, provision of additional measurements, and/or incorporation of prior information on decision variables. However, that does not guarantee that an improvement will be effective in converting an ill-posed problem into a well-posed one. This is done during the Inverse Problem Simplification process, which exploits prior information and knowledge of the system to simplify the *inverse problem*.
- The Parameter Sensitivity Analysis establishes a threshold detectable response value for each decision variable. This is done, following a comparison of the uncertainty in observations against the hydraulic impact caused by a simulated change in a model parameter, to exclude candidate parameters or parameter values that do not cause significant impact on the model hydraulics.
- The Search Space Optimisation uses an objective approach to reduce the search space before solving the *inverse problem*. This is done to provide a good starting point for the subsequent optimisation analysis, due to less solution combinations, and a reduced number of computations to explore the search space.

The next chapter presents methods that apply the proposed decision-support framework with the ultimate purpose to improve the accuracy of leak localisation and the model quality. To do this, the methods use head and flow measurements, and integrate a search space reduction technique to increase the reliability of the detected network parameters and speed up the optimisation process. Two practical simulation-optimisation methods are proposed, depending on the accuracy of the starting model.

CHAPTER 4

Methods for Leakage Detection and Localisation, and Model Calibration based on Search Space Reduction

4.1 Introduction

For an accurate leak localisation it is necessary to have a model that has been calibrated. The reason for this is that due to the uncertainty associated with an uncalibrated model, a large impact on pressure and flow is required to be able to identify the event. However, a calibrated model is not always available, thus, many small leaks that cause a local impact in the hydraulics of the system remain undiscovered. Leak localisation and model calibration are interlinked and dependent procedures (Figure 4.1). When a leak is found and the model is updated its quality improves, while when a model is calibrated its ability to detect and localise leaks also improves. This is because both procedures are undertaken with the aim to minimize discrepancies between the observations and modelled outputs. Therefore, more observations from sensors would enable utilities not only to identify accurately, quickly and economically unknown leaks that do not rise to the surface, but also to calibrate Water Distribution Network (WDN) models more effectively. This can minimize the overall leak run times and establish a foundation for improved model quality assurance.

Given the current status and trends in the availability of pressure and flow data, as well as in the complexity of *inverse problems* for WDN modelling, this chapter proposes two methods that both have an ultimate purpose of improved leak detection and localisation. This relies on having an initial model for leak localisation, i.e., whether it is calibrated or not. When a calibrated initial model is available, then leak localisation can be performed. The opposite situation is when the uncalibrated model has to be calibrated first before leak localisation is

performed. Based on the above, this Chapter proposes two practical simulation-optimisation methods:

(1) A Leakage Inspection Method (LIM), which highlights the leakage area and makes pinpointing of leaks faster.

(2) A Calibration Method (CM) which improves the WDN model accuracy so it can be used in (1).

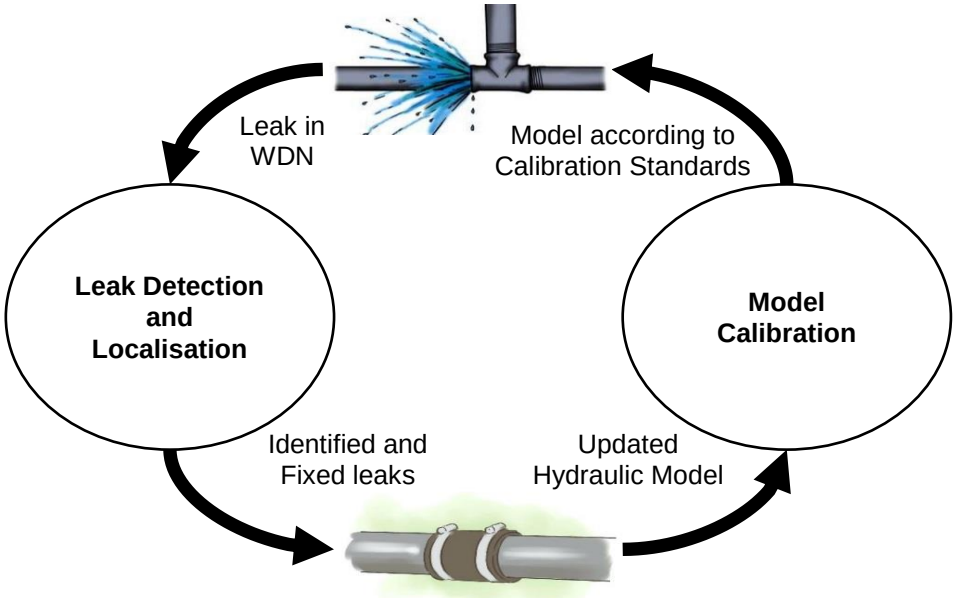

Figure 4.1. The leak localisation and model calibration feedback loop.

Both methods formulate the problem as an *inverse problem* and integrate a search space reduction technique (see Chapter 3) to reduce its dimensionality. The LIM requires a calibrated initial model of the WDN. On the other hand, the CM is implemented when a calibrated model is not available, so that it can serve as a basis for the LIM. However, the same series of uninterrupted observations is used in both methods. More specifically the LIM uses observations after a leak has happened and the CM (when used before the LIM) observations before a leak has taken place. Bearing in mind the above, this Chapter is organised as follows. After this introduction the philosophy of the LIM is presented in Section 4.2. Then, Section 4.3 provides an insight into the quality of monitoring, collecting and analysing data in WDNs that are used by both systems. Section 4.4 describes the methodological details of the LIM, while Section 4.5 addresses the CM. Finally, Section 4.6 provides a summary and conclusions for this chapter. Before proceeding with the description of the two methods it is meaningful to elucidate the rationale of presenting the LIM before the CM. Although an accurate

hydraulic model is necessary for carrying out leak detection and localization, the motivation for carrying out this research is the identification of leakages and not the overall improvement of hydraulic modelling.

4.2 The Overall Leakage Detection and Localisation Process

4.2.1 Introduction

There is no universally agreed procedure for finding leaks. Externally-based methods (see Section 2.5.3) are very accurate in pinpointing leaks on pipelines, however, the take long time in a large search space. On the other hand, internally-based methods (see Section 2.5.4), such as those that use hydraulic modelling, exhibit higher uncertainty in how accurately system and data faults can be identified. This is due to the fact that the WDN model is only a simplified representation of the real system, yet these methods require less effort than the externally-based methods. Though most frequently leak localisation is performed using externally-based methods, to expedite the leak localisation process a synergy of internally-based methods, such as WDN hydraulic modelling, with externally-based methods is needed.

The proposed data-driven leakage detection and localisation method exploits the advantages of internally-based methods to, eventually, facilitate externally-based methods for the ultimate purpose of faster and more accurate leakage pinpointing. The model-based search space reduction technique (see Chapter 3) can be used to reduce the number of computations required to explore the candidate leak locations. Therefore, through fast processing of the available pressure and flow data the LIM can highlight the areas of possible leakage. Consequently, this corresponds to a reduced search distance required for ground crews to pinpoint leaks, restricted to only part(s) of the WDN. To do this, an externally-based technique can be used, that is locally more accurate.

When a calibrated starting model is not available the ability to detect accurately and localise leaks reduces, due to the deteriorated model quality. Consequently, the model needs to be calibrated according to the some standards, so that is can be used for leak detection and localisation. The proposed CM follows the same

philosophy as the LIM for model calibration with the only difference being that the CM uses data before a leak has happened whereas the LIM after the leak. The historical, or incoming live measurements of pressure and flow provide information about the WDN state. Compared to an accurate model, an uncalibrated model may contain errors associated with the pipe, valve, or pump components. With sufficient data, such errors can be detected during the calibration process. In the proposed CM the *inverse problem* size is minimized through the implementation of a search space reduction scheme for pipe and valve components, before the model is calibrated.

The novel offline methodologies use heads (calculated from pressures and elevations) and flow field measurements either collected from planned field tests, i.e., when pressure sensors are installed in the District Metered Area (DMA) for a limited period, or communicated from permanently deployed loggers. The former would also mean that the run time of any hard-to-find leaks that occur after the calibration of the hydraulic model or the leakage campaign, cannot be minimized unless the next calibration period follows on after the previous one. In view of this, permanent sensor installation is recommended. Bearing this in mind, Figure 4.2 provides a schematic representation of the overall proposed leak detection and localisation procedure and data flow route.

It all starts from the pressure and flow monitoring devices, whereby the collected readings from each device are communicated and stored in the water utility's sensor database. A data pre-processing process is, then, applied to the raw set of measurements. Any corrupt or inaccurate measurements are detected and corrected (or removed) and, ultimately, a well-behaved (see Section 4.3.2) problem dataset is prepared. The problem dataset is, then, fed into the LIM, whereby the output result is a highlighted WDN section where a likely leak event has been identified. An externally-based method follows next searching for the exact leak location(s) within the area proposed by the internally-based method. Any identified leaks or changes in the state of the system are, then, updated within the model, which is also expected to become more accurate, i.e., better calibrated. The updated hydraulic model can be used along with new incoming data from the sensors, whereby the LIM can reassess the WDN for further faults.

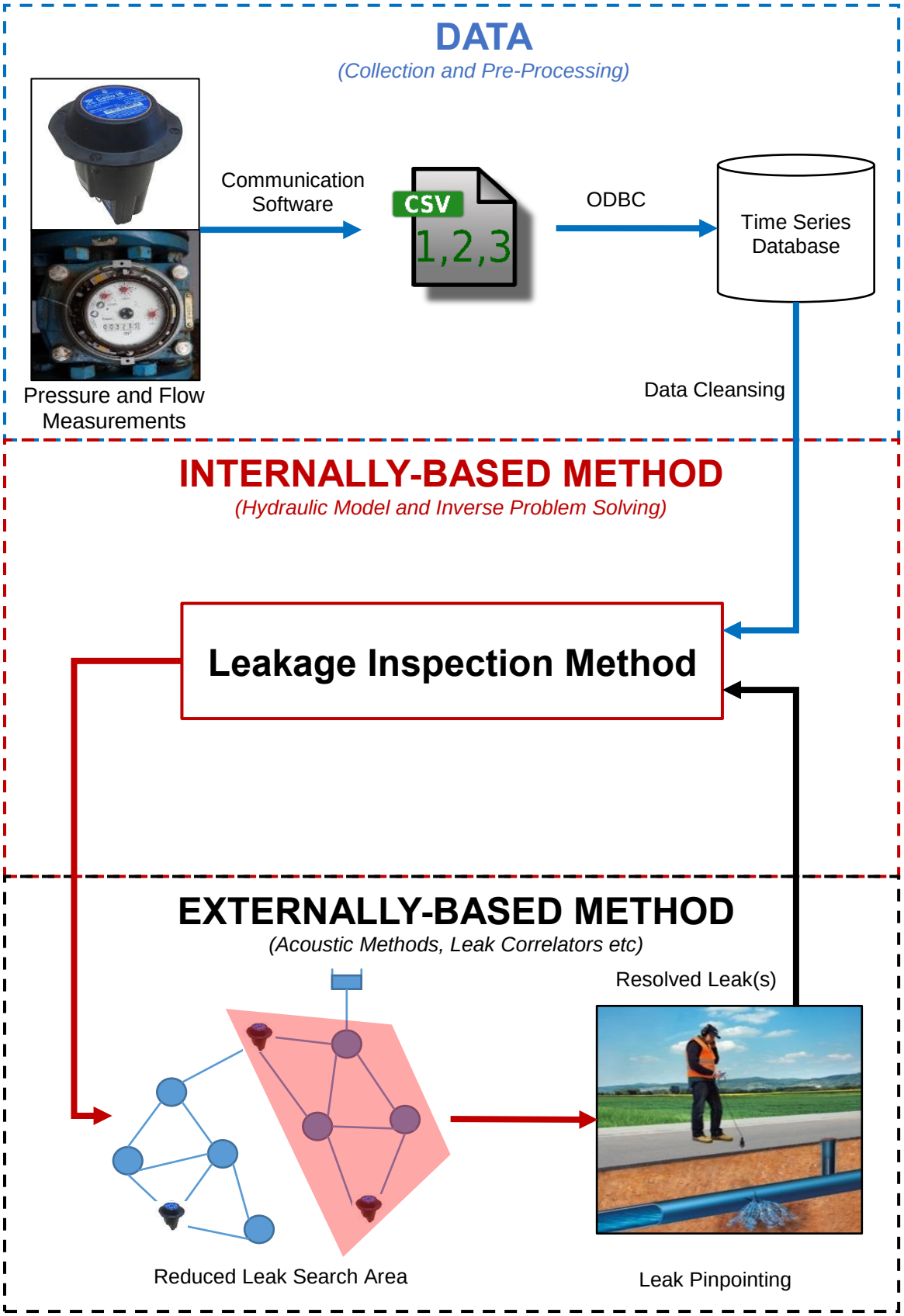

Figure 4.2. The proposed novel procedure used for leakage detection and localization.

4.2.2 Sensor Data

Measurement devices in a DMA are normally equipped with a Global System for Mobile communications (GSM), which is capable of individual General Packet Radio Service (GPRS) data communication (i.e., newest UK best practice for leading water companies). The use of GPRS allows the data collected at regular time intervals (e.g., 15 minute) to be communicated to the water companies without incurring a per-call charge that is characteristic of other communication technologies (e.g., GSM calls, Short Message Service – SMS messages). The data from each sensor may be communicated at longer intervals (e.g., every 30 minutes) to increase the sensor battery life. Thus, they typically capture readings every 15 minutes and initiate calls to the communication software, while in some cases the data from each sensor may be communicated at longer intervals in order to increase the sensor battery life. For each pressure and flow device a number of readings are obtained (e.g., 2 readings – assuming 15 minute sampled data, and communication frequency every 30 minutes).

An export functionality in the communication software is used to automatically update a discrete set of Comma Separated Values (CSV) files. An Open Database Connectivity (ODBC) text driver is used to interface to the Time Series database storing the time series history of the DMA signals. Along with pressures and flows, when a monitoring device is installed, the water company collects manual measurements of the location (i.e., latitude and longitude) and elevation (altitude) of the sensor using the Global Positioning System, as well as the depth from the ground level. These allow to determine heads from pressure signals. The collected raw data should be assessed so that any faults in the sensor devices are identified. The quality of the available datasets can be evaluated through a Data Pre-Processing scheme, where the resulting output is a set of equally separated measurements of pressures and flows (in terms of time interval) for a specified period useful for enabling the implementation of the LIM or CM.

4.2.3 Internally-Based Method

The data from all the DMA's sensors signals are, then, used within the LIM so that any leaks in the WDN are identified. Depending on the sensor type, number, location and accuracy there is a limited observable WDN space, i.e., the length of pipes that can be monitored for leakage. If the pressure perturbation from a leak or any other fault is below the device's accuracy range, the event will remain undetectable regardless of the distance from the sensor. However, this also allows to reduce the search space, which consequently decreases the computational and physical effort required to find leaks. Once the data from all the DMA signals are fully analysed within the LIM, the output of the leak detection and localisation methodology is a section of the network where a leak has likely occurred in the DMA. Additional information is also provided, which can be useful for the diagnosis of the event occurring, such as the path to pinpoint the event location. This provides a much shorter leak search distance, which enables the Leakage Technicians to pinpoint a leak faster, minimizing its run times.

4.2.4 Externally-Based Method

Once WDN section/part has been identified for leakage, an externally-based method is used to investigate the pipe section(s) in more detail and pinpoint the fault. Acoustic-based methods (e.g., listening sticks and leak noise correlators) still remain superior in terms of detection accuracy and the current-practice to sense a leak-induced signal. However, further developments of other equipment-based leak detection methods are envisaged (e.g., pig-mounted acoustic sensing devices and/or ground penetrating radars). Pinpointing and fixing the fault does not complete the process, as after a leak is fixed the hydraulic model needs to be updated. This leads to additional benefits associated with model calibration. Based on a new set of measurements the WDN can be reassessed, which could lead to the identification of further leaks in it. It is, therefore, suggested that the process is carried out iteratively.

4.3 Sensor Data Pre-Processing

4.3.1 Introduction

A well-known characteristic of sensor data is that it can be uncertain and erroneous. The reason for this is that sensor batteries can fail, the communication network might experience interruptions, and/or the tendency of sensors to wear or drift. Other factors, such as low-cost sensors, freezing or heating of the casing or measurement device, accumulation of dirt, mechanical failure or vandalism (from humans or animals), affect heavily the quality of the sensor data. In general, these sources of errors can be classified broadly as either systematic errors (bias) or random errors (noise). Systematic errors arise due to changes in the operating conditions, e.g., temperature, humidity, etc., or other factors such as ageing of the sensor. The effect of random errors on sensor measurements may seriously affect the accuracy of the WDN modelling analyses. The sources of random errors include, but are not limited to:

(1) Noise from external sources;
(2) Random hardware noise;
(3) Inaccuracies in the measurement technique (i.e., readings are not close enough to the actual value of the measured phenomenon);
(4) Various environmental effects and noise; and
(5) Imprecision in computing a derived value from the underlying measurements (i.e., sensors are not consistent in measuring the same phenomenon under the same conditions).

All these errors may cause a significant problem with respect to data utilization, since leak detection and localisation using erroneous data may yield unsound results. Additional errors in the results can be caused by faulty elevation or depth measurements, as these are used to calculate heads. Hence, data quality assurance is a vital step prior to any analysis or application. The quality of decisions taken will depend on the quality of data used in the WDN modelling analyses. To address this problem, it is essential to detect and correct erroneous values in sensor data by employing data pre-processing. This activity converts raw data into fit-for-use data without errors, duplicates, and inconsistencies, i.e., high-quality data.

4.3.2 Sensor Data Quality

Readings obtained from WDN monitoring devices can, in general, be classified as “dirty”, i.e., missing or unreliable (Ediriweera & Marshall, 2010). The main contributing factors for dirty readings is the lack of operator knowledge of communication systems issues and sensor management, weaknesses in the specification given to the equipment vendor, back end Information Technology system failures, and failing transducer hardware. Poor data quality manifests as sections of missing data, data from faulty loggers (offsets in reading values), and the presence of erroneous date stamps (offsets in reading times). Typical well-behaved data (Batini & Scannapieco, 2006) from flow and pressure sensors should be characterised by:

- **Accuracy:** It indicates the extent to which data reflects the real system hydraulics. If it is not 100% that means that there are errors in data.
- **Completeness:** It refers to whether all available data is present, i.e., there are no gaps in the datasets. When data is due to unavailability, this does not represent a lack of completeness.
- **Consistency:** The data is kept up to date and agrees with itself. This means there is no difference, when comparing two or more representations of the sensor data.
- **Timeliness:** The degree to which data represent reality from the required point in time.
- **Uniqueness:** Entities are recorded once and there should be no data duplicates reported.
- **Validity:** Data are valid if it conforms to the syntax (format, type, range) of its definition.

It is important to obtain up-to-date diagnostic information about the devices and conduct data validation to ensure the accuracy and validity of subsequent analysis or application. Communication failure alarms for remote sites are often overlooked, and the collection of increasing amounts of data requires a schedule for data validation. It has been estimated that water utilities in the UK use only 10% of data (Water Briefing, 2013; Karimova, 2016), due in large part to low confidence in data quality. Ideally, basic data should be validated each time a logger is downloaded. Data checking should include:

(1) Whether the connection and download was successful.
(2) Whether the data is correctly stored in the data management system.
(3) The presence or absence of data.
(4) The presence of zero or negative values, which usually indicate logger/sensor failure.
(5) Whether the data value and time is “reasonable”.

Some telemetry software systems provide facilities for flagging some of these issues. For example, alarm limits may help detect zero values. Validation techniques may be employed to highlight failing data points through exception reports and thus facilitate improved maintenance. However, often some manual interpretation is required from staff. As stated earlier, different locations within a network can yield significantly different levels of information about system performance. A further concern relates to the accuracy of monitoring devices used during field testing. Pressure and flow monitoring devices are accurate within a specified range. Provided that equipment calibration is maintained, standard sensors used for modelling purposes are normally accurate within a range of ±0.1% (Halaczkiewicz & Klima, 2018) of full scale 0-10 bar (i.e., typically accurate to ±0.1m), while flow meters within a range of ±1% (Peterson, 2018) for the recorded reading. Advances in pressure sensor calibration technology have now made it is possible to achieve accuracy ranges of ±0.0185% (General Electric, 2018), which is a 5-fold improvement in data quality. Pressure data should be collected at strategic points such as low pressure regions, control structures, areas of water quality complaints, and areas providing good calibration data. Significant improvements in the accuracy of the model calibration and leak localisation process can be secured with the inclusion of flow measurements captured from key flow routes in the WDN. Hence, data collection should be focused on such areas.

4.3.3 Data Pre-Processing

Data Pre-Processing (DPP) is the process of detecting and correcting (or removing) corrupt or inaccurate records from the sensor reading data (Rao et al., 2012). This is achieved by identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or

coarse data. DPP is an activity that consists of implementing error prevention strategies, such as screening, diagnosing, and editing (Van Den Broeck et al., 2005) during:

(1) Data Collection.
(2) Data Communication/Transfer.
(3) Data Transformation/Extraction.
(4) Data Analysis/Exploration.

Screening involves looking systematically for suspect features in the raw dataset (ACAPS, 2016). When screening data, it is convenient to distinguish oddities, such as lack or excess of data, outliers including inconsistencies and strange patterns. The identification of an error is followed by *diagnosis*, i.e., finding the cause for the defective data. The diagnosis and *treatment* phases of DPP requires an in depth understanding of all types and sources of possible errors encountered during data collection and entry processes (ACAPS, 2016). During diagnosis the treatment for the problematic observations can be:

(1) **Left unchanged:** The most conservative course of action is to accept the data as a valid response and make no change to it. However, in a small dataset the impact of such action on the analysis result will be larger.
(2) **Edit:** Changing the value of data shown to be incorrect. On the other hand, a change based on engineering judgement may bias the data as a result of introducing subjective impact.
(3) **Deleted:** The data seems illogical and the value is so far from the norm that it will affect leak localisation or model calibration process. The drawback is that less data will be available.

Following treatment, a set of pressure or flow data should be consistent with other similar data sets in the system and is now ready to load into the leakage inspection system. This output data from DPP is of high-quality, i.e., standardized, uniform, accurate and complete and can expedite the processing speed and performance of overall leak detection and localisation system. The DPP approach is demonstrated in Figure 4.3.

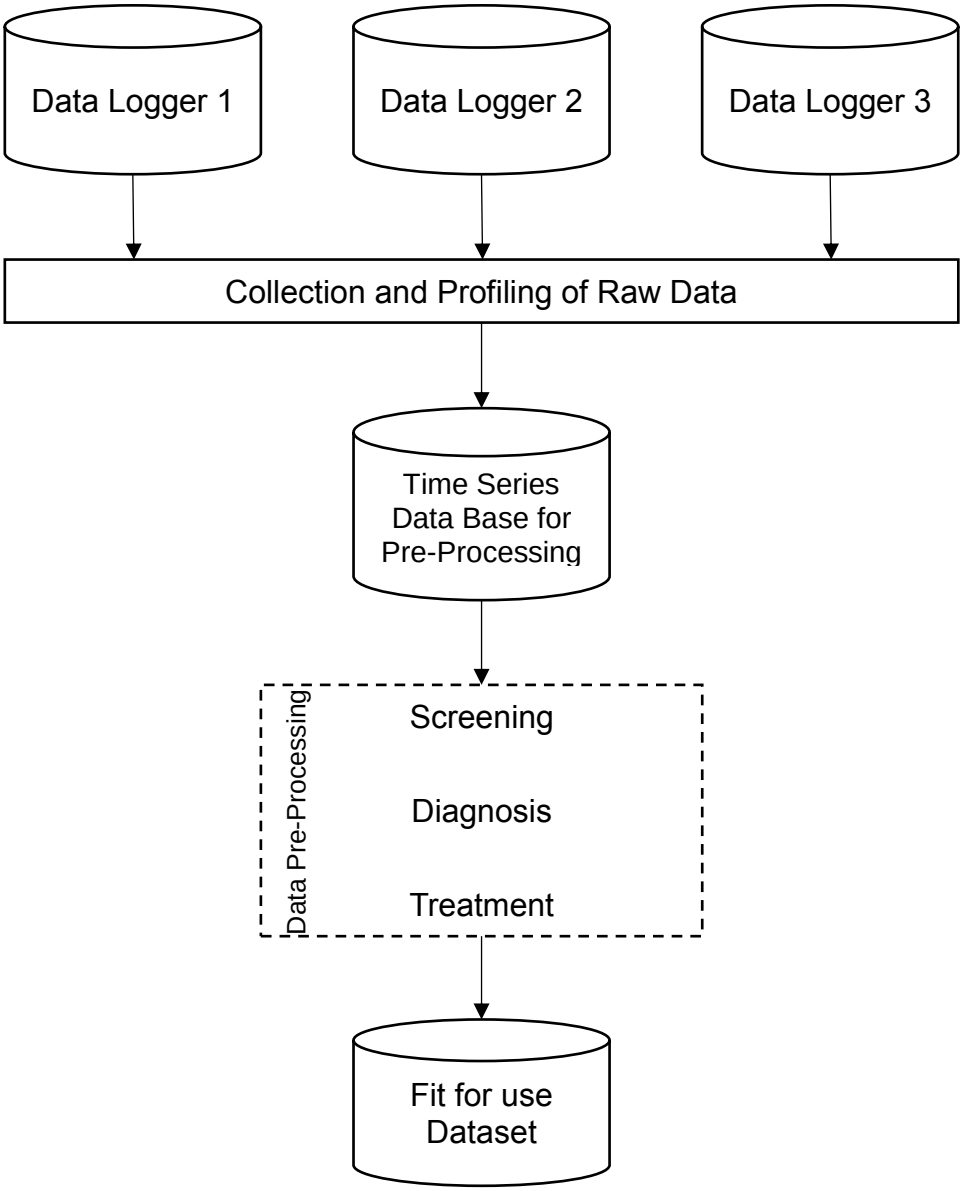

Figure 4.3. Schematic Representation of Data Pre-Processing.

One serious data quality issue that must be addressed on the collected measurements prior to leakage detection and calibration analysis is that of missing readings. If the quantity of available data is sufficiently large, and the size of the affected readings is small, then the simplest approach may be to discard those patterns from the dataset (Batini & Scannapieco, 2006). This approach is based on the assumption that the omission of the data values is independent of the data itself and the description of patterns. If this is not the case, then this solution will modify the data distribution in an adverse way. Such an example would be a sensor that always fails to produce an output signal when the signal value exceeds a particular threshold.

When there is “too little” data to consider (based on engineering judgement) discarding potentially useful information contained in partial patterns, then techniques for “filling in” such data must be considered (Dasu & Johnson, 2003). For a continuous stream of data such as a time series, it will usually be preferable to fill in the data to recreate the continuous stream. Various heuristics may be adopted for dealing with this “filling in”. The simplest approach is interpolation from adjacent points (which is only useful for small numbers of missing values). Alternatively, each missing value can be replaced by the mean of the corresponding variable over those patterns for which a value is available. For time series with fairly predictable cyclic behaviour, this technique is reasonable and can be used in a seasonal context. A more sophisticated approach is to express any variable that has missing values in terms of a regression over the other variables (assuming they are correlated) using the available data, and then use the regression function to fill in the missing values. To some extent, the method adopted also depends on whether isolated missing values are considered or there are long sequences of missing values, for which a more sophisticated approach may prove necessary. Bearing this in mind, note that this topic is beyond the scope of the work presented here and hence will not be discussed in greater detail.

4.4 Leakage Inspection Method

4.4.1 Overview

The Leakage Inspection Method (LIM) detects and localises leaks in DMAs, based on systematic search space reduction. The leak localisation is formulated as an *inverse problem* where unknown leaks are simulated as emitter coefficients representing leak flow discharge. The leak location(s) and size are determined using an optimisation technique, where decision variables include the candidate leak locations and their corresponding emitter coefficients. A practical simulation-optimisation framework highlights the leakage area and makes pinpointing of leaks faster. The quality of a generated solution is evaluated through the simultaneous comparison of the available heads and flows measured by deployed sensors with the simulated values from the hydraulic model. The LIM

method integrates two stages (Figure 4.4) based on the developed decision-support framework for solving *inverse problems* proposed in Section 3.2:

1. a Search Space Reduction (SSR) stage for reducing the *inverse problem* size;
2. a Leak Detection and Localisation (LDL) stage for finding leaks.

All simulations in this research were carried out using the EPANET Programmer's Toolkit (Rossman, 2000), while MATLAB was used to implement the developed optimisation-based algorithm. Eventually, a computer-based tool has been developed for the LIM, which automates the simulation-optimisation process and eliminates any manual task by the user on the hydraulic model. The full description of the automated tool is given in Appendix A, as the main focus of the thesis is on the developed methodology and not the software. During the SSR stage the LIM initially leverages the space reduction technique for node components proposed in Chapter 3. This is so that the search domain before leak localisation is objectively reduced and unnecessary simulation of solutions that cause no impact on model fitness is avoided. This is done by implementing three search reduction steps (see section 3.2) considering the WDN model nodes. Then, at the LDL stage an optimisation problem is solved, to indicate the size and location of leaks in the WDN. A calibrated hydraulic model is a prerequisite of this approach. A schematic framework of the LIM is given in Figure 4.4.

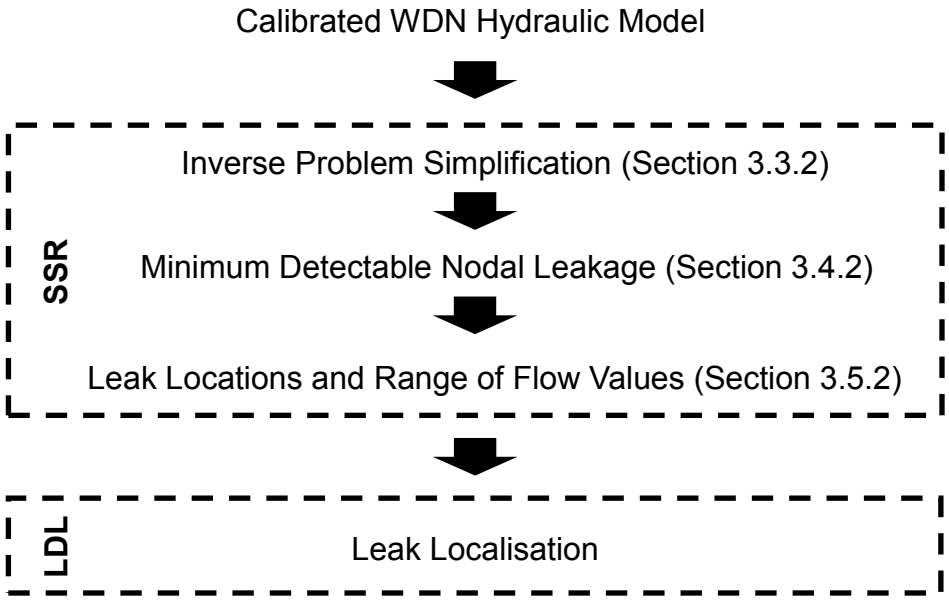

Figure 4.4. Overview of the Leak Localisation methodology framework.

4.4.2 Assumptions

The effectiveness of the developed method relies on the following key assumptions:

- Flow measurements from the inlet and outlet meter(s), as well as heads from pressure measurements at an indicative minimum spatial resolution of at least one sensor per 200 properties, are available (this is the current standard employed by the water utility supporting this work). This is collected at a time interval of 15 minutes and a duration of 24 hours to allow a full Extended Period Simulation (EPS) analysis.
- Any leaks identified by the proposed method are located at network nodes. Note that this does not reduce the generality of this method and can still detect leaks located along network pipes.
- The hydraulic model of the analysed water system is calibrated according to the acceptable standards (Ormsbee & Lingireddy, 1997) prior to any leakage detection.
- The impact of any existing error in calibration, such as the existence of valves with unknown status or pipe friction coefficients, on the modelled outputs following leak detection and localisation is minimal.
- The leak flow rate remains unchanged in the time horizon of the data used to detect and localise leaks.

4.4.3 Search Space Reduction Stage

The WDN model nodes are considered and the three-step search space reduction method proposed in Chapter 3 is implemented to minimize the number of candidate leakage node locations and the range of possible flow values. The outcome of the SSR stage establishes a reduced set of decision variables, x_{SSR}, for solving the *inverse problem*. The Inverse Problem Simplification at Step 1 removes any nodes related to non-pipe components (see section 3.3.2). Then, at Parameter Sensitivity Analysis at Step 2 the Minimum Detectable Nodal Leakage for each candidate location is determined, where a user input is necessary for the starting flow value (see section 3.4.2). Based on the number, location and accuracy range of sensors a minimum bound flow for each location is established. Finally, the Search Space Optimisation at Step 3 defines the maximum number

of possible leak locations in the WDN and ultimately the range of possible flow values for each potential location (see section 3.5.2). The output from the SSR stage is a reduced list of candidate leak locations and range of possible flow values that is considered as the basis of the LDL stage, where the size and location of leaks is identified.

4.4.4 Leakage Detection and Localisation Stage

At the LDL stage, another *inverse problem* is formulated, similarly to Step 3 of SSR stage (see section 3.5.2), to detect and localise the emitter flow values in the WDN. The optimisation searches for a maximum of $NLeak$ possible leaks based on the fittest n scenario resulting from Part II of the Search Space Optimisation step. The decision variables include the nodes remaining in the list of leak locations J and the range of their flow values, following Part II. The range of c_i^n values, involves a vector of equally separated discretized values between the upper, $\overline{c^n}$, and lower, $\underline{c^n}$, bounds established in Step 3 of SSR, also including the value of zero for no leakage. Again, the flow and head observations are considered during the optimisation analysis to minimize Equation 3.16 (see Section 3.5.2). Eventually, in a reduced set of decision variables, the leak detection and localisation process starts with a more targeted possible solutions as opposed to a case where no reduction is applied.

4.5 Calibration Method

4.5.1 Overview

The Calibration Method (CM) determines the state of internal pipe roughness values, detects the status of any throttle valves with uncertain position, and identifies the setting/speed of any pressure reducing valve/variable speed pump, as well as the multiplier coefficient of any demand pattern. Therefore, any network component that has an impact on the pressure and flow value and profile is calibrated. More specifically, the adjustment of the pipe roughness and throttle valve status can affect the WDN model pressure, while the demand pattern multiplier coefficients have an impact on flow. The optimisation of Pressure Reducing Valves (PRVs) and variable speed pumps is associated with the simulation of the response of those network components according to changes

in demand. The calibration is formulated as an *inverse problem* and applies the systematic search space reduction technique proposed in Chapter 3. Its aim is to produce a hydraulic model that accurately simulates the average day hydraulic conditions of the real system and hence can be used for leak detection and localisation. Through the use of head and flow data and an optimisation technique the aim is to improve the quality of the hydraulic model, i.e., to correctly simulate the flow path head losses and directions, as well as pressure and flow profiles in the WDN. In addition to speed up the calibration process and reduce the search effort required to identify errors associated with unknown closed or open throttle valves. The quality of a generated solution is evaluated through the simultaneous comparison of the available heads and flows measured by deployed sensors with the simulated values from the hydraulic model. A novel DPP approach is used to generate the problem dataset from the raw sensor data comprised of measurements that represent the average day conditions.

The Calibration Method involves four stages (Figure 4.5):

(1) Data Pre-Processing (DPP) stage to generate the calibration dataset.
(2) A Profile Calibration (PC) stage for macro-level calibration of the WDN model.
(3) A Search Space Reduction (SSR) stage for reducing the calibration problem size.
(4) A Component Calibration (CC) stage for micro-level calibration.

Similarly to the LIM, the EPANET Programmer's Toolkit and MATLAB was used to implement the simulation-optimisation framework, leading to the development of a similar computer-based tool for automating all the stages in CM. The full description of the CM tool is provided in Appendix B following the main emphasis of this research. In addition, similar assumptions were made associated with the density and the interval of the collected flow and head measurements. On the other hand, the necessary duration of measurements is at least two weekdays to allow the representation of average day conditions. Furthermore, the impact of any existing error in calibration, such as the existence of background leaks, on the modelled outputs following model calibration is assumed to be minimal or null.

In the DPP stage the head and flow observations are pre-processed to convert the raw set of measurements to a 24-hour dataset that represents the average day hydraulic behaviour of the WDN. This is performed by implementing two steps:

(1) Average Day Profile Dataset Generation.

(2) De-Noised Dataset Generation.

Following the DPP stage the problem dataset comprised of measurements for heads and flows for calibration is constructed, which is then used as an input for the next stages. During the PC stage the profile of any demand category and of any model component's setting that affects the WDN model flow and pressure, respectively, is calibrated using inverse analysis. Each profile is developed using pattern multiplier coefficients applied against a baseline value for each time step. The GA optimises the profiles by implementing two steps, where decision variables include the pattern multipliers for candidate demand category or model component's setting, respectively. The two steps include:

(1) Demand Profile Calibration (DPC).

(2) Pressure Profile Calibration (PPC).

At the SSR stage, the pipe and valve components are considered and the search domain is reduced before the calibration of the WDN model, by implementing the three proposed steps (see Chapter 3). The decision variables include the pipe group roughness coefficients and the status of candidate valves. The three steps of SSR are implemented for each decision variable type separately and the outcome is a reduced list of candidate pipe roughness groups and valve components, which is considered in the CC stage. Then, at the CC stage, another optimisation problem is solved where the pipe group roughness coefficients and status of valves are fine-tuned to calibrate the WDN model at micro-level and indicate any unknown closed or open valves in the WDN. The updated hydraulic model should be calibrated according to the performance criteria and be available for uses in leak detection and localisation. A schematic framework that demonstrates the four stages of the CM is given in Figure 4.5.

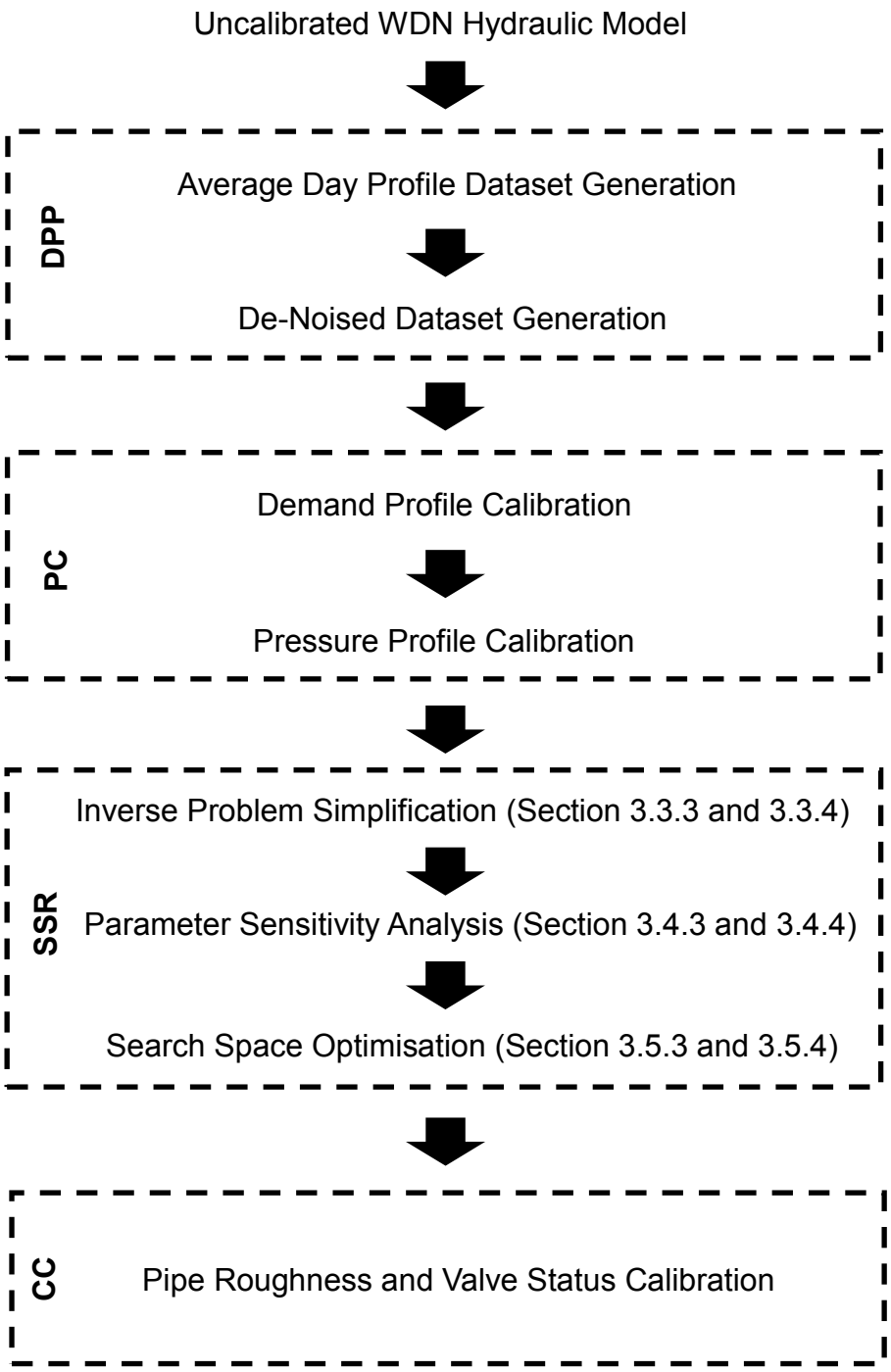

Figure 4.5. Overview of the Calibration Methodology framework.

4.5.2 Data Pre-Processing Stage

4.5.2.1 Background

Water usage in WDNs is inherently variable due to continuously changing demands. In order for an EPS model to reflect accurately the hydraulic behaviour of the real system, these demand fluctuations must be incorporated as profiles.

The temporal variations in the water consumption of WDNs typically follow a 24-hour cycle, called a diurnal demand profile, which also produces a corresponding pressure profile due to head losses. However, system flows experience changes not only on a daily basis, but also weekly, annually and in a long term. Seasonal differences in water usage have been related to climatic variables, such as temperature and precipitation, as well as the changing habits of customers, such as outdoor recreational and agricultural activities occurring in the summer months. As one might expect, weekend usage patterns often differ from weekday patterns, due to changing habits.

The EPS hydraulic models are typically calibrated with the aim to produce an accurate representation of the average day's pressure and flow profile in the real WDN, associated with a weekday. However, in practice, choosing the best dataset to calibrate the WDN model is subjective, as it depends on the engineer's judgement and experience of the system. In a calibration campaign, the monitoring devices are typically deployed for limited period of time recording data for approximately 2-3 weeks. However, during the calibration process observations from one day (24 hours) are chosen out of all the recorded data. The choice for the day that represents the calibration dataset is based on the engineer's judgement and experience of the WDN, so that it represents the hydraulic behaviour of an average day as closely as possible. Normally, the calibration dataset is associated with measurements recorded on either a Tuesday, Wednesday or Thursday. The rationale for this is because Monday and Friday are closer to weekend days and, thus, the consumption pattern may be affected by the weekend demand profile. On the other hand, the choice of a single day (either Tuesday, Wednesday or Thursday) as the calibration dataset for the average day conditions in the real system introduces a lot of uncertainty, due to the day-specific variability in consumption.

4.5.2.2 Overview

The Data Pre-Processing stage develops an objective method for preparing the calibration data with the aim to improve the calibration of average day models. The method prerequisites well-behaved data characterised by accuracy, completeness, consistency, timeliness, uniqueness and validity. The developed

Data Pre-Processing method uses the head and flow observations that are recorded during weekdays, i.e., Monday to Friday, to create an objective and accurate dataset for the average day hydraulic behaviour of the WDN. In other words, it aims to create the best reflection of observed data used to calibrate the hydraulic model. This is done in two steps:

(1) Average Day Profile Dataset Generation

(2) De-Noised Dataset Generation.

In Step 1 (4.5.2.3), the pressure and flow data from weekdays (without weekends) are considered to generate a dataset of the average day for each monitoring device, where the raw set measurements is converted to a 24-hour dataset. Then, in Step 2 (Section 4.5.2.4) a data de-noising technique is implemented on the average day dataset to remove any noise caused by day-specific outliers. The resulting set of smoothed and equally separated head and flow measurements for 24-hours is then, used by the CM as the calibration dataset. An example of how the 24-hour calibration dataset is generated from raw observations during the DPP stage, is presented in Figure 4.6.

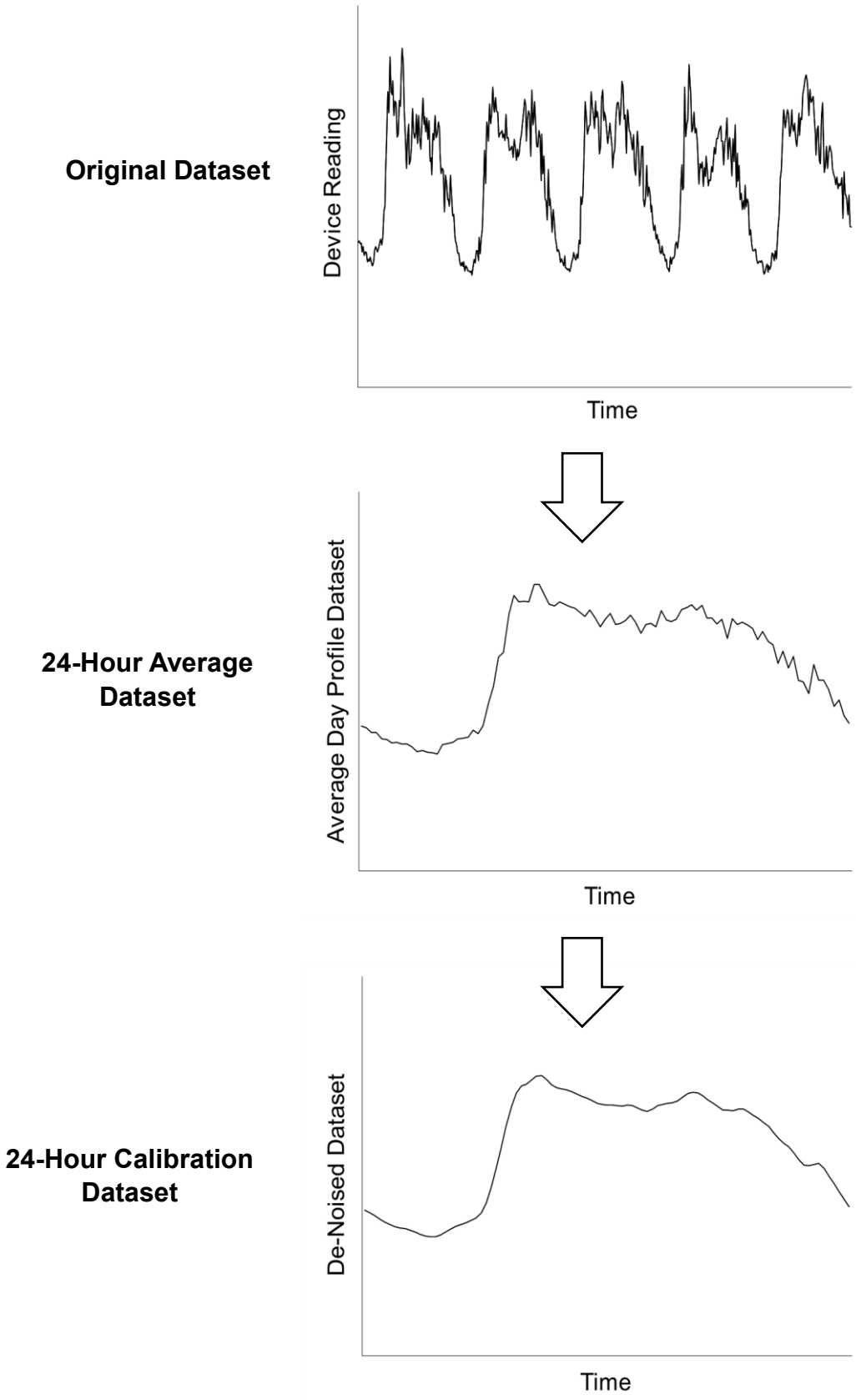

Figure 4.6. Example of the data pre-processing method applied on the available flow measurements.

4.5.2.3 Average Day Profile Dataset Generation

In the Step 1 of the DPP stage the raw dataset of weekday heads from pressures and flow measurements from each device is used to create a 24-hour dataset that represents the average day. The number of time steps in the average day dataset depends on the temporal resolution of the available measurements. The increment readings captured by each device are used to determine the average value for head and flow at each time step, using:

$$\overline{y_{o,m}}(t) = \frac{\sum_{d=1}^{T} y_{o,m,d}(t)}{T} \quad (4.1)$$

Where: $\overline{y_{o,m}}(t)$ is the average value of the readings from device m at time t;
$y_{o,m,d}(t)$ is the measurement value from device m at time t on day d;
T is the total number of days in considered in the dataset of measurements.

Ultimately, a 24-hour set of measurements is created for each device, representing the average day profile for the period of recorded data. An advantage of this method is that it does not require each device to have the same number of measurements, i.e., the total number of days, T, as it is based on the average conditions.

4.5.2.4 De-Noised Dataset Generation

Still, however, due to outliers in measurements caused by day- and time-specific variability in consumption, the resulting dataset pattern contains noise and may not be representative of the realistic average day conditions. For example, if the consumption pattern of a major customer has abruptly changed for a specific day when observations were recorded, this may significantly affect the average day dataset. To avoid such an impact on the calibration dataset as a result of temporal variability in consumption patterns, in Step 2 a smoothing technique is implemented to de-noise ADPD of each device.

Smoothing techniques are often employed to capture general patterns in stressor-response relationships among variables and eliminate noisy data. The

classical procedure to smooth a dataset is to fit a polynomial of suitable degree. The problem with polynomial smoothing is that it is not resistant. A few data points at the extreme right of the dataset can very much affect the fitted values at the left of the scatterplot. In fact, this is a general problem of fitting curves with the least squares method. To tackle this, the **Lo**cally **We**ighted Polynomial Regre**ss**ion (LOWESS) method is used which is a data analysis technique for producing a "smooth" set of values from a time series which has been contaminated with noise, or from a scatter plot with a "noisy" relationship between the 2 variables. It was introduced in the statistical literature in the late 1970s (Cleveland, 1979) and later developed by many others (e.g. Jacoby, 2000; Loader, 2004).

LOWESS regression is based on a smoothing procedure which pays greater attention to the local points. The smoothed value y_o^* corresponding to observation $y_{o,k}$ is obtained on the basis of the observations around it (nearest neighbours) within a band of certain width (i.e., time steps). The LOWESS procedure determines each smoothed observation by computing the regression weights for observation in the band. The smoothing band can vary in size and is specified as a fraction between 0 and 1, which represents the percentage of data (0-100%) the band covers. The point $y_{o,k}$ is the midpoint of the band. The total data points, P, within the band are assigned weights in a way so that $y_{o,k}$ has the highest weight. The weights for the other data points decline with their distance from $y_{o,k}$ (local weighting), according to a weight function, given by the tri-cube function:

$$w_k(t) = \left(1 - \left|\frac{y_{o,i} - y_{o,k}}{d_i}\right|^3\right)^3 \tag{4.2}$$

Where: w_k is the is the tri-cube weight function for observation $y_{o,k}$ at time step t;

$y_{o,k}$ is the observation value to be smoothed;

$y_{o,i}$ is the nearest neighbour, i of $y_{o,k}$ as defined by the span;

d_i is the time distance from $y_{o,k}$ to its P^{th} observation (the most distant value) within the span.

The weights have the following characteristics:

- The data point to be smoothed has the largest weight and the most influence on the fit.
- Data points outside the span have zero weight and no influence on the fit.

A weighted linear least-squares regression is performed using a first degree polynomial. The smoothed value is given by the weighted regression at the predictor value of interest. The procedure is repeated for all the data points in the 24-hour dataset and for each device. The outcome is a De-Noised dataset of generated measurements for 24 hours corresponding to the average conditions for each measurement device at each time step. The generated dataset is used as the observations by the CM so that the WDN model is calibrated to represent the average conditions for the period of recorded data.

4.5.3 Profile Calibration Stage

4.5.3.1 Background

The pressures and flows in a hydraulic model are simulated as temporal patterns based on the demand type associated with each WDN model junction at a specific time step. When more than one demand type is allocated to a particular junction, the demand is said to be a composite. Determining the total rate of consumption for a junction node with a composite demand is a simple matter of summing the individual demand types. When temporal patterns are applied to composite demands, the total demand for a junction at any given time is equal to the sum of each baseline demand times its respective pattern multiplier for each time step of the hydraulic simulation. The series of demand pattern multipliers models the diurnal variation in demand and can be reused at nodes with similar usage characteristics. Therefore, any adjustment in the series of multipliers can directly affect flow and, consequently pressure. The baseline demand is often chosen to be the average daily demand (although peak day demand or some other value can be used).

It is generally accepted in WDN modelling that demand patterns repeat every 24 hours with only negligible differences. The amount of time between

measurements has a direct correlation to the resolution and precision of the constructed profile, i.e., the number of demand pattern multipliers. The demand profile for each junction can be constructed given by:

$$Q_j(t) = \sum_{p=1}^{P} B_{j,p} QPat_{j,p}(t) \quad (4.3)$$

Where: $Q_j(t)$ is the total demand at junction j at time t;
$B_{j,p}$ is the baseline demand for demand type p at junction j;
$QPat_{j,p}(t)$ is the pattern multiplier for demand type p at junction j at time t.

Nevertheless, flow and pressure are inversely related. An increase in consumption causes more head loss and, thus, reduction in pressure. On the other hand, the pressure of the system can be controlled at any time through the use of PRVs that reduce the available head, or pumps that increase it. For example, when a DMA is pressure managed the PRV setting may be set at a lower level during the night for leakage reduction purposes. Therefore, the pressure profile for each WDN part can be constructed by:

$$P_k(t) = BS_k PPat_k(t) \quad (4.4)$$

Where: $P_k(t)$ is the pressure at the model component k at time t. The model component k can be either a source node (e.g., reservoir or tank), a PRV or a pump, where the head pattern, setting or speed can be adjusted, respectively.
BS_k is the baseline head, setting or speed at model component k;
$PPat_k(t)$ is the pattern multiplier for the baseline head, setting or speed BS_k of model component k at time t.

4.5.3.2 Overview

The Profile Calibration is divided into two parts where different optimisation problems are solved with the ultimate purpose to macro-calibrate the WDN hydraulic model. During this stage, the model parameters that are associated with

the diurnal variation in demand and pressure profile are calibrated. Therefore, changes are made to the values of the PRV settings, pump speeds and the demand pattern multiplier coefficients. The calibration is formulated as an *inverse problem* where decision variables include the discrete multipliers for any pattern that impacts either the demand or pressure in the WDN model. Using the generated 24-hour calibration dataset after the DPP stage the aim is to optimise the discretized profile of any demand category and any setting of any model components associated with the hydraulic behaviour of the WDN. In both parts, the problem is formulated to optimise the pattern multiplier coefficient of one time step at a time, so that discrepancies between observed and simulated outputs are minimized. The number of pattern multipliers is equal to the total number of time steps associated with the problem dataset. The process begins by optimising the pattern multiplier for the first time-step of the EPS (e.g., at 00:00 hours) and ends when optimisation of all pattern multipliers for the length of the hydraulic simulation has been completed. The two parts of PC stage involve:

(1) Demand Profile Calibration (DPC)

(2) Pressure Profile Calibration (PPC).

During DPC the decision variables include the pattern multipliers for all non-leakage demand categories associated with the hydraulics of the WDN model and where measurements are available. Using only observed flow measurements from the generated calibration dataset the aim is to calibrate the profile of each demand type and consequently the model's overall demand at all entry points. DPC is carried out before PPC as demands drive the head losses in the system. At the end of DPC, the resulting modelled flows should exactly match the available observations. Then, in PPC a second optimisation problem is solved with the updated demand profiles in the hydraulic model, to calibrate the profile, setting or speed of any model component that impacts pressure in the WDN. For example, the speed of a pump can be adjusted as a response to a change in WDN demand in order to maintain a desired pressure. Consequently, the optimisation of those model components aims to simulate such a situation. In this part only heads from pressure observations are considered and the outcome from this process is a hydraulic model that is calibrated at a macro-level and serves the foundation for the CC stage. The optimisation problems are solved similarly

in both parts and are subject to two different sets of constraints: (1) the set of implicit type constraints considering system hydraulics; and (2) the set of explicit constraints used as bounds for the algorithm solution search space for each decision variable. The optimisation in each part is formulated as:

Search for: $\vec{X} = (PM_t^p) \quad PM_t^p \in P^p; p = 1, \dots, Npattern; t = 1, \dots . NStep$ (4.5)

Minimize: $F(\vec{X})$ (4.6)

Subject to: $[\underline{PM_t^p} \leq PM_t^p \leq \overline{PM_t^p}]$ (4.7)

Where: PM_t^p is the pattern multiplier for pattern p at time step t;
P^p is the set of pattern multipliers for pattern p;
$Npattern$ is the total number of patterns to be calibrated;
$NStep$ is the number of time steps for each pattern p.

4.5.3.3 Demand Profile Calibration

In DPC part a series of short (user specified) optimisation analyses, $NStep$, are undertaken, where at each analysis, t, the demand multiplier, PM_t^p (equivalent to the multiplier $QPat_{j,p}(t)$ in Equation 4.3), is optimised for $Npattern$ demand types with the aim to minimize the weighted sum of squared flow difference given by Equation 3.13. Any demand type associated with the leakage profile is not considered in this process as it is assumed that the background leakage remains constant. Following all optimisation analyses, the outcome is a hydraulic model with calibrated demands that represent the average day consumption pattern.

4.5.3.4 Pressure Profile Calibration

A similar optimisation problem is formulated in PPC part using the updated hydraulic model. During PPC, $NStep$ optimisation analyses are undertaken, where at each analysis, t, the pattern, setting or speed multiplier, PM_t^p (equivalent to the multiplier $PPat_k(t)$ in Equation 4.4), is optimised for $Npattern$ components. These are associated with the drive head of any source node, the setting of any PRV, or the speed of any pump in the WDN, respectively. The optimisation problem minimizes the weighted sum of squared head difference given by Equation 3.22. The outcome of PPC part is a hydraulic model where simulated

heads match the pattern shape and are close to observed ones, but still do not match them due to additional errors in the WDN model. The hydraulic model following the PC stage is considered for performing the SSR stage, where the micro-level calibration problem is minimized.

4.5.4 Search Space Reduction Stage

The WDN model pipes and valves are considered and the three step search space reduction method proposed in Chapter 3 is implemented to minimize the number of candidate pipe groups and valves for roughness and status calibration, respectively. The three steps of SSR are implemented for each decision variable type separately, i.e., one decision variable type is considered at a time. The Inverse Problem Simplification at Step 1 removes any valve of known or estimated status and allocates the WDN model pipes into groups according their characteristics (see section 3.3.4 and 3.3.5). Then, at the Parameter Sensitivity Analysis at Step 2 the Detectable Pipe Roughness Coefficients and Valve Components are determined, where the number, location and accuracy range of sensors is taken into account (see 3.4.4 and 3.4.5). Finally, the Search Space Optimisation at Step 3 defines the maximum number of pipe calibration groups and possible closed valves in the WDN (see 3.5.4 and 3.5.5). Following the SSR stage a reduced list of candidate pipe roughness groups and valve components, is prepared and considered as the basis of the CC stage, which is described in the next sub-section.

4.5.5 Component Calibration Stage

At the CC stage, another *inverse problem* is formulated to calibrate the hydraulic model at a micro-level. The aim of this stage is to improve the model quality and identify any undiscovered errors carried out from previous calibration procedures, which are associated with incorrect pipe states and flow topology. An optimisation problem is formulated to where decision parameters include the reduced set of pipe roughness groups and candidate valves with unknown status following the SSR stage. The optimisation searches for a maximum of $NGroups$ possible calibration groups and $NValve$ possible closed valves based on the fittest scenario resulting from Step 3 of SSR stage for each decision variable type. The decision variables include the groups of pipes and the range of their possible

roughness values, along with the valves remaining in $RV_{pg,pm}$ and in the list of valve locations, V, respectively, following Part II. The range of λ_j^g values for each candidate pipe group, involves a vector of equally separated discretized values between the upper, $\overline{\lambda_j^g}$, and lower, $\underline{\lambda_j^g}$, bounds based on $DPRC_{pg,pm}$. Again, the flow and head observations are considered during the optimisation analysis to minimize Equation 3.16. The optimisation is formulated as:

Search for: $\vec{X} = (\lambda_j^g, CV_k^v)$; $\lambda_j^g \in RV_{pg,pm}$; $CV_k^s \in V^s$

$g = 1, \dots, NGroups$; $v = 1, \dots, NValve$

$j = 1, \dots, RIndex$; $k = 0, \dots, VIndex$ (4.8)

Minimize: $F(\vec{X})$ (4.9)

Subject to: $\underline{\lambda_j^g} \leq \lambda_j^g \leq \overline{\lambda_j^g}$ $P > 0$ (4.10)

Where: λ_j^g is the roughness coefficient value for the group of pipes g corresponding to index j in set $RV_{pg,pm}$ associated with pg of material class pm;

CV_k^s is the index for valve k for the possible closed valve s;

$RV_{pg,pm}$ is the set of possible roughness values that belongs in $DPRC_{pg,pm}$ for the group g associated with pg and material class pm;

V^s is the set of candidate valves for the possible closed valve s;

$NGroups$ is the total number of pipe groups to be calibrated;

$NValve$ is the number of possible closed valves to be identified;

$RIndex$ is the number of possible roughness values for pipe group, g, associated with pg and material class, pm;

$VIndex^s$ is the number of the candidate valves for the possible closed valve s,

$\underline{\lambda_j^g}$ is the minimum and $\overline{\lambda_j^g}$ is the maximum roughness value associated with pipe group g.

P is the head at any node.

4.6 Summary and Conclusions

The head and flow data from deployed sensors provide a potentially useful source of information about the state of the WDN, which is beneficial for procedures such as leak detection and localisation, as well as model calibration. A novel offline model-based method for leak detection and localisation that makes use of these data has been presented in this chapter. This method is implemented in a computer-based system and has been developed as two separate tools, based on the quality of the starting model. The LIM, highlights the leakage area and makes pinpointing of leaks faster, whereas the CM improves the WDN model quality so it can be used for more accurate leak detection and localisation.

The main conclusions from this chapter are as follows:

- The accuracy of leak localisation and the effectiveness of model calibration is affected by the performance of each method, i.e., the LIM and CM, respectively. For an accurate leak localisation a calibrated model is necessary, as its ability to identify leak events improves with more accurate predictions. Similarly the hydraulic model predictive quality improves when a leak is found and the model is updated after the leak is fixed.
- Internally- and Externally-based methods for leak detection and localisation have their individual advantages and disadvantages. A synergy of both methods is necessary for faster and more accurate leakage pinpointing.
- The developed LIM and CM make use of a hydraulic simulation model of the WDN and an optimisation algorithm with the main purpose to enable more effective and efficient discovery of leaks and mimicking of the behaviour of the WDN. Both methods are implemented on several case studies, which are demonstrated in the next chapter.
- The quality of the simulated outputs depends on the quality of data used in the WDN modelling analyses. A data pre-processing method has been proposed for detecting and correcting erroneous values in sensor data. Its abilities in converting raw data into fit-for-use data are shown in the next chapter

- The LIM uses a practical simulation-optimisation framework and systematic search space reduction to detect and localise leaks in DMAs. This allows to highlight the leakage area and expedite leak pinpointing.
- The CM determines the state of internal pipe roughness, the setting/speed of any PRV/pump and detects the status of valves with uncertain position.
- The SSR stage integrated in both systems can minimize the *inverse problem* dimensionality to only a number of highly sensitive decision variables, depending on the sensor type, number, location and accuracy. This can speed up the computational performance of the LIM and CM.
- The novel DPP method serves as an objective approach for preparing the calibration dataset that captures the best reflection of general hydraulic patterns of the WDN in an average day. Its abilities in contributing to an improvement in model quality, and consequently leak localisation are demonstrated in Chapter 5.

In the next chapter the LIM and CM methods are applied to both semi-real and real case studies and the results from each case study are presented and discussed, along with an evaluation of the proposed methods.

CHAPTER 5

Application of the Leakage Inspection and Calibration Methods to Case Studies

5.1 Introduction

This Chapter presents and discusses the results of four case studies where the Leak Inspection Method (LIM) and Calibration Method (CM), proposed in Chapter 4 are validated and demonstrated. Both artificially generated and real pressure and flow data are used on semi-real (SR) and real (R) case studies, respectively, from Water Distribution Networks (WDNs) at District Metered Area (DMA) level in the United Kingdom (UK). In all presented case studies the ultimate purpose is the same, i.e., to detect and localise leak events. An overview of all the presented case studies is given in Table 5.1. The main objectives of these analyses are as follows:

- To determine the benefits of the search space reduction stage in speeding up the *inverse problem* solving procedure.
- To demonstrate and evaluate the capabilities of the LIM in increasing the reliability of detecting (i.e., reducing false positives) and localising (i.e., reducing the distance from the actual anomaly) leaks in both semi-real and real cases.
- To assess the effectiveness of the CM in improving the WDN hydraulic model quality so it is fit for the purpose of leak detection and localisation.
- To compare various calibration datasets under different hydraulic conditions and assess the corresponding model along with the overall effects on the leak localisation accuracy.

Table 5.1. Overview of the presented Case Study Applications.

Case Study	Network	Data	Model Calibration	Rationale	Leak Detection and Localisation	Rationale	Test Conditions
SR1	Real	Synthetic	No	-	Yes	Understand LIM robustness in finding leaks	A range of leak sizes and sensor reading accuracy values
SR2	Real	Synthetic	Yes	Understand the improvement in starting model accuracy using the CM	Yes	Understand effect of calibration on leak localisation and LIM robustness under demand uncertainty	Calibration before leak localisation, using different datasets and for a range of demand uncertainty scenarios
R1	Real	Real	Yes	Calibrate model using the CM	Yes	Validate LIM effectiveness on a real leak event	A calibrated hydraulic model of the WDN is not available at the time of the leak
R2	Real	Real	Yes	Calibrate model using the CM	Yes	Validate LIM effectiveness on a real leak event	A calibrated hydraulic model of the WDN is not available at the time of the leak

The SR1 case study aims to test the robustness of the LIM in finding leaks. To do this the LIM is implemented for a range of leak sizes and under a range of sensor reading accuracy values. In the SR2 case study the LIM is implemented to localise a number of leaks after firstly calibrating the hydraulic model based on various scenarios. The same model is calibrated based on four possible datasets and under a range of demand uncertainty scenarios. The aim is to assess the effectiveness of the CM in improving the starting model before leak localisation and investigate how the calibration procedure affects the accuracy of leak localisation. Also, to understand the effects of demand fluctuation in finding leaks. In the R1 and R2 case studies, the CM is firstly implemented to calibrate the WDN hydraulic model of a real system. Then, the LIM is used to detect and localise a real leak event that happened using field data captured from deployed flow and pressure devices. The aim is to validate the methodologies proposed for the LIM and CM on two different real systems.

The chapter is organised as follows. After this introduction, Section 5.2 reports the results obtained after applying the novel systems to the semi-real case studies SR1 and SR2. Then, in Section 5.3 the results of the real district water system examples obtained in the real case studies R1 and R2, are presented. In each of these sections the description of the case study begins with details of: (a) the studied area; (b) the data used for the analyses; and (c) the problem setup in the specific case study being analysed. This introduction to the problem is, then, followed by the results of the various optimisation analyses that were performed. Once this is done, Section 5.4 discusses the operational and computational benefits of the two systems and evaluates their effectiveness, efficiency and practicality. Finally, a summary of the chapter and the main conclusions are given in Section 5.5.

5.2 Semi-Real Applications

5.2.1 Case Study SR1

5.2.1.1 Experimental Design

The aim of this experiment is to determine the robustness of the leak detection and localisation methodology and establish the minimum leak flow that could be

detected and localised. A series of desktop type experiments were conducted to detect and localise a leak of varying size for a range of sensor reading accuracy values using the LIM (Table 5.1). A number of scenarios based on a single leak of varying size were created where the locations of the leak and sensor devices remained constant. The impact of the timing of the leak was not investigated in this research and, thus, was assumed at 00:00 so that a full 24 hours of data were available. An artificial set of pressure and flow 'observations' was generated for each scenario using an Extended Period Simulation (EPS) analysis in EPANET considering the system state containing a leak. A number of nodes and the inlet pipe were chosen as locations for monitoring pressure and flow, respectively. This was determined based on the current standard in pressure sensor deployment by the water utility supporting this work, where 1 logger is installed for every approximately 200 households. The artificial measurements were generated for leak sizes of 50%, 33%, 25%, 20%, 15%, 10% and 5% relative to the average network inflow, here named as flow scenarios. Therefore, the capabilities of the LIM were tested for both small (5% and 10%), medium (20% and 25%) and large (33% and 50%) leak sizes. Four datasets were created for each leak flow size, resulting in 6 x 4 = 24 scenarios. The first includes perfect measurements, whereas the other datasets correspond to pressure measurements with random noise of various magnitude. The noise was generated considering a uniform distribution resulting in values ranging between ±0.1%, ±0.25 and ±1% of the reading, here named as noise levels. The choice for the tested noise values was based on the accuracy of current pressure monitoring devices used in field testing, allowing for an additional 10-fold increase in noise (Halaczkiewicz & Klima, 2018).

5.2.1.2 WDN Hydraulic Model Description

The SR1 example is based on a real urban DMA in the UK, which is highly looped and is fed via a single inlet. The WDN model (Figure 5.1) is composed of 202 junction nodes and 158 pipes with the total length of 9.4km. In addition, there are 59 valves which are not presented in Figure 5.1. The average WDN pressure is 25.3m with a maximum difference of 6.9m across the WDN. During a leak-free situation, flow from the source node varies between 1.3 l/s at Minimum Night Flow (MNF) and 7.8 l/s at morning peak demand, with an average of 3.6 l/s. The DMA

has flow and pressure sensors installed at the inlet and seven inner pressure sensors. The placement of the pressure sensors is marked using red circle symbols in Figure 5.1. 'Field measurements' generated through EPS simulation were recorded every 15 minutes for a period of 24 hours. This provided the total of 96 simulated field data points - from midnight to midnight. Each data set represents a complete snapshot of the system conditions. The observed inflow into the DMA and at the pressure at eight pressure sensor locations were used for evaluating the quality of leak detection and localisation.

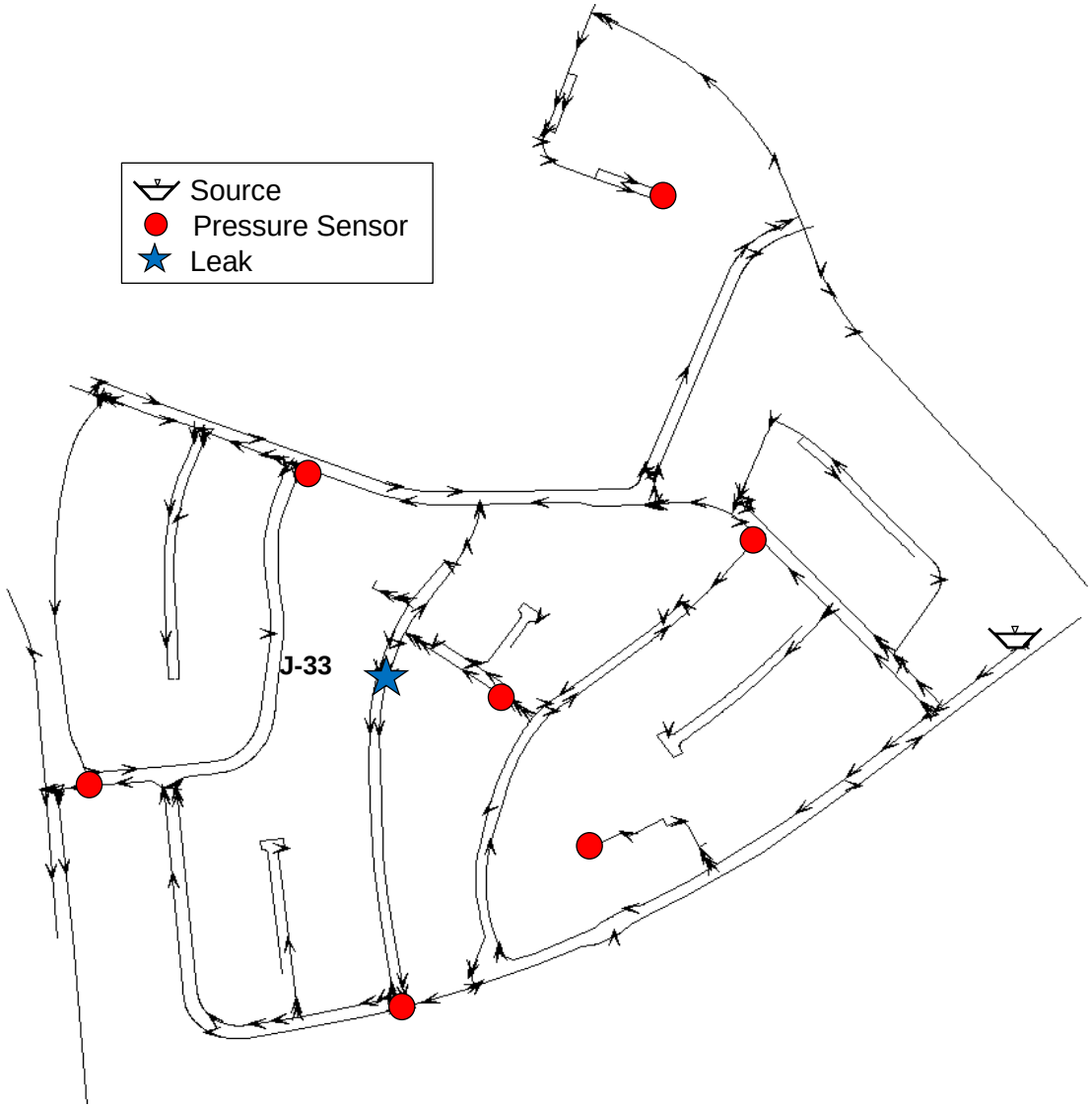

Figure 5.1. The SR1 Case District Water System.

5.2.1.3 Leak Localisation Problem Set Up

A leak was introduced at J-33, shown by the blue star symbol in Figure 5.1. The location at J-33 was chosen to occur on a looped part of the WDN. Looped WDN

parts are more hydraulically reliable than branched WDN parts. Thus, there is less hydraulic impact from a leak event, which makes it more difficult to identify this leak. The pressure loggers were placed both upstream and downstream of the leak location. This was expected to allow the impact of the leak to be observed in most of the pressure data. The method was tested for leak sizes of 3.6 l/s (50%), 1.8 l/s (33%), 1.2 l/s (25%), 0.9 l/s (20%), 0.65 l/s (15%) 0.4 l/s (10%) and 0.2 l/s (5%). Furthermore, four observed datasets were generated for each flow scenario. The first of the four sets of experiments used perfect pressure and flow data (i.e., with no noise) for all flow scenarios. The subsequent datasets involved a systematic pressure error ranging between ±0.025m, ±0.063m and ±0.250m, respectively, per recorded time step and for each device. On the other hand, the typical error range found in flow meter devices of ±1% (Peterson, 2018), was introduced to the flow measurements of all flow scenarios.

5.2.1.4 Overall Inverse Problem Reduction

The starting model contained no leakages. Based on the noise level, the Minimum Detectable Nodal Leakage (MDNL) for each candidate leak location differed. The MDNL process lasted approximately 10 minutes for each tested scenario. This led to a distinct range of possible flow values for each candidate node. Hence, following Step 3 of SSR, which took about 30 minutes to complete, this resulted in a different number of candidate nodes for each flow scenario. At the end of SSR, Part I, the detected water losses for each leak scenario always equalled the size of the actual simulated leak. Furthermore, any node with higher MDNL value was correctly excluded from consideration. Following on from Part II and under all flow scenarios, the fittest optimisation analysis corresponded to the WDN with a maximum one leak. Thus, in all cases the Leak Detection and Localisation (LDL) part involved two decision variables, i.e., the location and the size of a single leak.

The percentage of nodes that remained in the final list of candidates for each flow scenario and for each noise level is presented in Figure 5.2. The base case used the perfect data and involved 42% of the WDN nodes. This resulted in 84 out of the 202 nodes as the potential locations for the single leak across all flow scenarios. The reduction was achieved by *inverse problem* simplification. As

noise level increased (or the leak size decreased), the number of candidate nodes was reduced. This is due to the fact that less accurate data provide a smaller flow resolution range for each node, i.e., possible leak flow values between the maximum and minimum emitter coefficient, because of a higher MDNL value for each location. Consequently, there is less chance of finding leaks when the level of uncertainty is similar or higher than the head loss caused by the event. The downside of this is that leaks causing a small hydraulic impact cannot be detected. Interestingly, for a leak size of 5% (0.2 l/s) and a noise level of ±1% (±0.25m) the methodology fails to identify any candidate leak nodes for this WDN. The low head loss changes caused by the small leak, combined with the high pressure sensor noise level results in an undetectable leak situation. Therefore, the approach fails to detect and localise a leak for that flow scenario and the given noise level. However, in all other tested cases the optimum solution J-33 always remained in the reduced search space used for LDL.

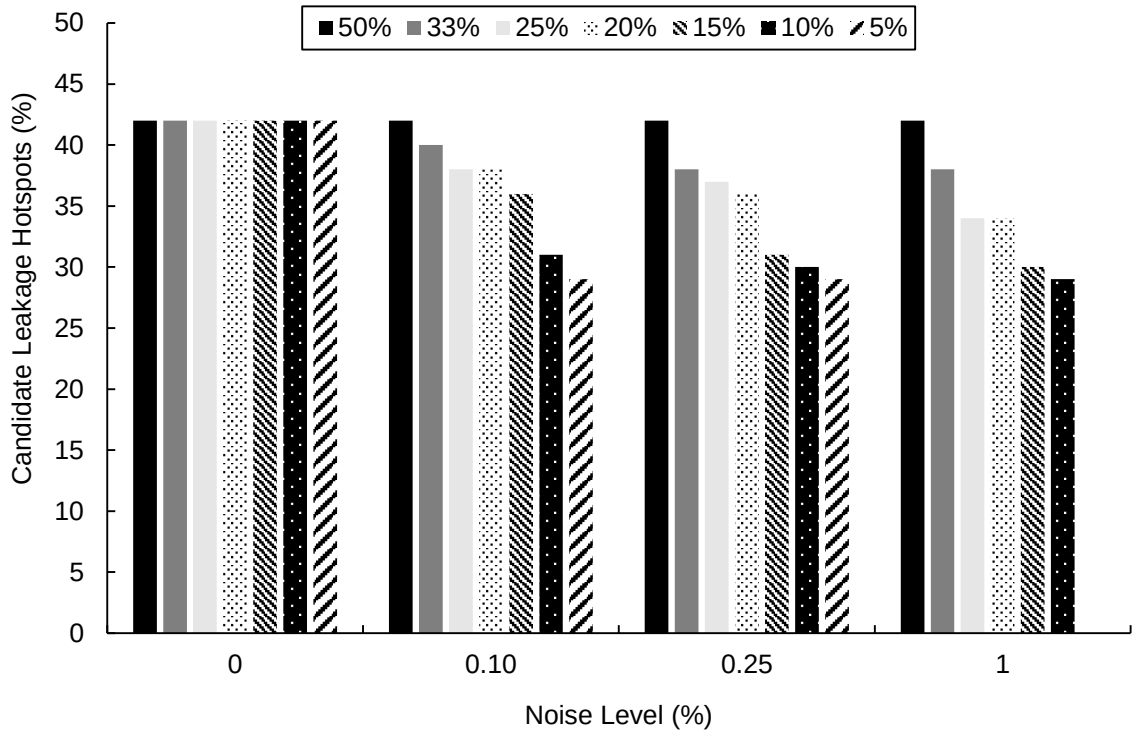

Figure 5.2. Percentage of the total WDN nodes representing potential leak locations that are left for the use with the detection and localisation methodology based on different leak sizes (figure legend) and noise levels.

5.2.1.5 Leak Detectability

The mean pressure differences between 'observations' and simulated outputs for each leak scenario and noise level, are shown in Figure 5.3. The numbers given are shown as a percentage change relative to the leak-free situation. The data labels illustrate the corresponding pressure drops in metres of water head.

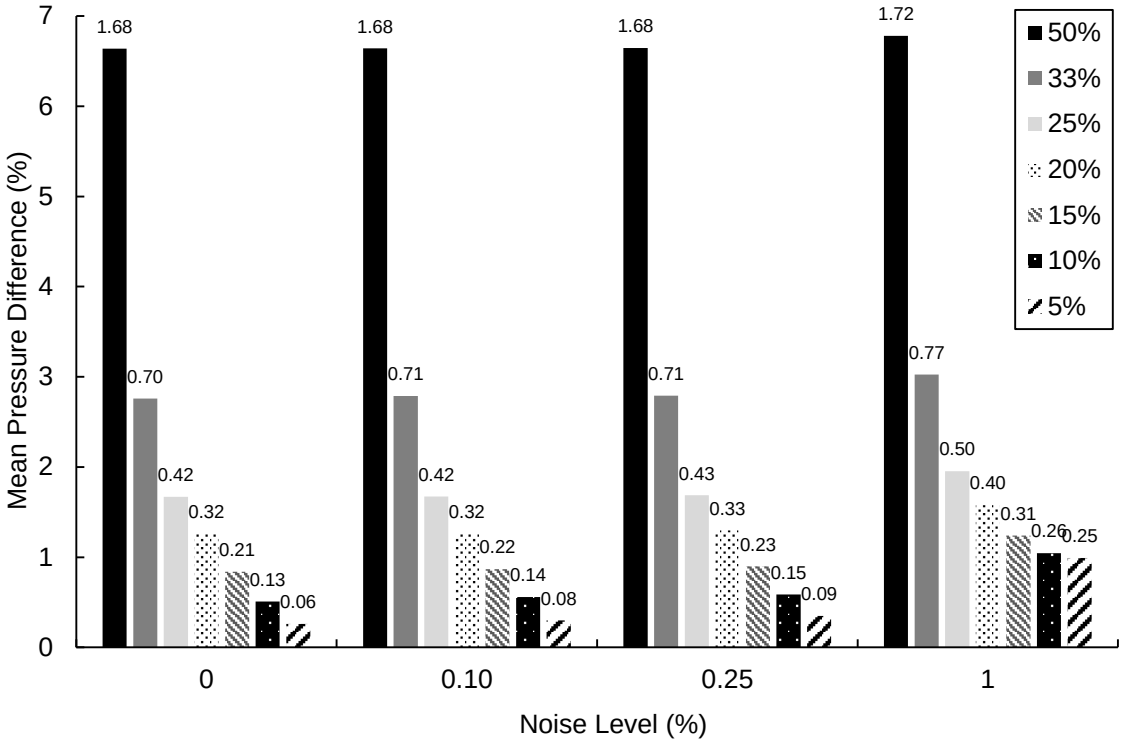

Figure 5.3. Mean Percentage Pressure Difference for each leak flow scenario relative to leak-free case.

As it can be seen from Figure 5.3, the mean pressure difference from the leak-free case decreases with a decreasing leak size. Relative to the base case, the difference in pressure changes caused by different flow scenarios decrease with a reduced leak size and an increasing noise level. Interestingly, the pressure change at a noise level of 1% (±0.250m) is similar for leak sizes of 5% and 10%. Compared to the mean pressure perturbation, noise has a small impact at larger flows. At a 50% leak scenario there is a pressure drop of 1.68m, while this changes to 1.72m at a noise level of 1%, which comprises around 2.5% of the perturbation. In contrast, a 5% leak only causes 0.06m pressure change, using perfect dataset, whereas at the typical noise level of ±0.1%, the noise accounts for 25% of the variation. This variation rises to 75% at 1% noise level. Under such

condition, it would be impossible to localise a leak due to the large uncertainty. Furthermore, the small difference between the pressure changes at lower flows can result in a non-unique problem. As the leak size reduces, the observable change in the pressure data also decreases. This makes it difficult to find a distinct “signature” defined by the leak location J-33.

Table 5.2. Comparison of geographical distances (m) of the detected leak from the assumed (true) leak location for each flow scenario and noise level.

Noise level (%)	0.00	0.10	0.25	1
Leak Size (%)	Distance between the detected and actual leak locations (m)			
50	0	0	0	0
33	0	0	0	0
25	0	0	0	0
20	0	0	0	0
15	0	0	0	0
10	0	9	9	420
5	42	171	171	N/A

Table 5.1 presents the distances of the leaks identified during the LDL stage relative to the true leak location at J-33. The LDL stage for each scenario was completed within one hour. As it can be seen from this table, the leak areas were successfully detected and localised up to the size of 10% (0.4 l/s) and up to the noise level of 0.25% (±0.063m). The latter corresponds to a mean pressure drop of 0.15m, where noise accounts for approximately 40% of the variation (Figure 5.3). This demonstrates that it is possible to detect and localise a leak if the contribution on the pressure drop is larger than the variation caused by the introduced noise (simulating the accuracy of the device). Even with perfect data at a flow of 5%, the pressure impact is small, i.e., less than ±0.3%. This makes it difficult to identify a unique pressure signature for J-33 relative to the node detected 42m away. The approach fails to detect any leak at ±1% noise level and a leak size of up to 10%. This is due to the noise impact being larger than the pressure perturbation from the leak. In this particular case and assuming the noise level of contemporary logging devices, the approach can narrow down the

search area for finding a leak of the size up to 5% of the inlet flow. In this example, this corresponds to an actual leak search distance range within ±2% of the WDN, considering the mains length. This shows that the methodology can find the area of the leak even with considerable noise in the data. From a water utility point of view, the methodology can reduce the leak search distance range in a real WDN. By reducing it to a street level, the approach can contribute to earlier pinpointing of the leak. At a larger noise level, the uncertainty is so high that the recorded pressure variation was mainly caused by the reading error and not by the leak itself. This suggests that more loggers would be required for improved leak localisation.

5.2.2 Case Study SR2

5.2.2.1 Experimental Design

In SR2 a series of desktop experiments (summarised in Table 5.3) were carried out to create a calibrated WDN model that can serve as a starting point for leak detection and localisation. The aim was two-fold. Firstly, to assess the effectiveness of the CM in improving the accuracy of average day models used to localise leaks (Table 5.1). Secondly, to compare different datasets used for model calibration and their effect on leak localisation accuracy. At each scenario, the starting WDN hydraulic model was calibrated and then used to determine the size and location of some leaks. Similarly to SR1, a number of nodes and the inlet pipe were chosen as locations for monitoring pressure and flow, respectively. Eventually, the leak localisation accuracy of each experiment was used to draw conclusions on the performance of the calibration methodology and compare the different datasets.

A number of Demand Fluctuation (DF) scenarios were initially created, whereby the system demand was varied differently at each time step of the day compared to other days. This was to emulate a situation whereby a unique consumption pattern is produced every day, depending on the DMA consumption types. Urban and rural DMAs are being emulated together with domestic and industrial consumption types. For example, due to the unpredictable water usage coming from less users, the demand changes more abruptly in a rural DMA, as opposed

to a more stable pattern in an urban area. The same stands for a domestic type consumption over the more constant industrial consumption pattern. The flow chart for the experimental design process is presented in Figure 5.4.

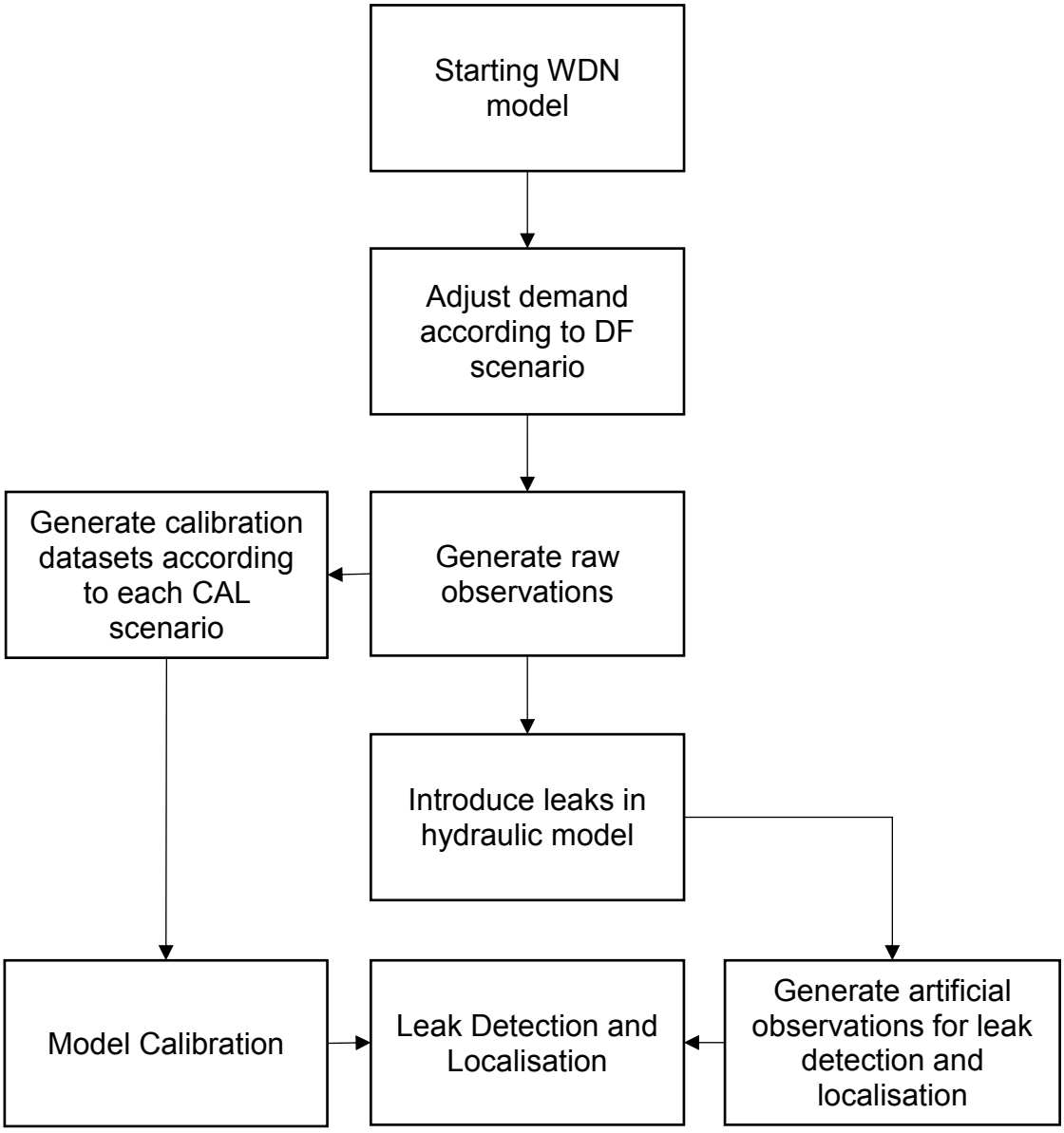

Figure 5.4. The general process flow chart for the experimental design.

A number of DF scenarios were created. The first scenario (DF1) assumed an ideal condition where no variability exists, i.e. every day has the same demand profile. The rest of scenarios were generated so that the demand at a specific time step of each day can vary within a ±2.5% (DF2), ±5% (DF3), ±7.5% (DF4) and ±10% (DF5), compared to the average value of that time step. The variability

was generated based on a random noise following a uniform distribution. Then, artificial 'observations' were generated for each DF scenario following an EPANET EPS analysis. Five days (i.e., 120 hour simulation) of "raw" observations, i.e., sensor data without any processing, were generated considering the system state with the actual pipe roughness values and status of valves. The five days of data represents a realistic field data collection period when deployed sensors record measurements for a specified period (e.g., one week).

The collected raw 'observations', were, then, pre-processed before being used as the 'calibration dataset' in the model calibration procedure. All the artificial 'observations' were generated to only emulate weekday consumption so that the calibration accuracy is not affected by the different weekend consumption pattern. Furthermore, an average weekday involves the system hydraulic conditions which the calibration outcome aims to simulate.

For each DF scenario a different dataset was chosen to calibrate the model. The following Calibration (CAL) scenarios were tested:
CAL A: A single day of data was chosen as the calibration dataset from the five day "raw" observations, i.e., similar to the traditional calibration dataset.
CAL B: A single day of data was chosen from the five day dataset, where the "raw" observations were smoothed to generate a de-noised calibration dataset.
CAL C: All raw observations were used to generate a calibration dataset corresponding to the average day.
CAL D: All raw observations were used to generate a de-noised calibration dataset of the average day, i.e., following the DPP stage.

Eventually, this resulted in 5 x 4 = 20 scenarios for calibrating the WDN model which are presented in Table 5.3.

A number of leaks were introduced at randomly chosen locations of the WDN model to create the dataset for leak detection and localisation. Again, a dataset of artificial pressure and flow measurements were generated for each DF scenario, as explained in section 5.2.2.1. A 24 hour dataset was then used for

the purpose of leak detection and localisation so that a full EPS analysis can be performed. This is not representative of a real situation, where water utilities investigate any unexpected differences during the night to search for a leak. However, as the consumption in the WDN is low during the night, the leak impact on pressure can be insufficient, compared to the impact on flow, which can be used for leak detection. Therefore a 24 hour dataset provides more information on the leak pressure impact over the whole day which can be used for localising leaks. All generated datasets corresponded to measurements with random noise following using a uniform distribution. This resulted in a pressure value range of ±0.1%, of the reading (typical error found in modern pressure measurement devices) and flow value range of ±1%.

5.2.2.2 WDN Hydraulic Model Description

The SR2 example is also based on a real urban DMA network in the UK, which is highly looped. The system, which is fed by one inlet, is outlined in Figure 5.5. The WDN hydraulic model is composed of 589 junction nodes, 178 valves (which are not shown in the presented layout) and 433 pipes with the total length of 14.6km. Out of the total number, 58% of pipes (250) are made from Cast Iron (CI), 20% (87 pipes) from Asbestos Cement (AC) and 22% (96 pipes) from a High Performance Polyethylene (HPPE) material. The average WDN pressure is 43.8m with a total range of 48m (21m – 69m) across parts of the WDN. The DMA has flow and pressure sensors at the inlet, and 11 inner pressure sensors, emulating a true situation where the sensor deployment density of 1 per 200 properties, whose placement is also marked using red circle symbols in Figure 5.5. The system is gravity-fed, whereby during a leak-free situation flow from the source node varies between 4.5 l/s at MNF and 18.9 l/s at morning peak demand, with an average of 11.1 l/s (without any leakages). Generated field measurements for calibration purposes were assumed to be recorded every 15 minutes. The total period corresponded to 120 hours (five days) or 480 simulated field data. On the other hand, 96 signals for 24 hours were produced for leak localisation purposes, i.e., from midnight to midnight. Each data set represents a complete snapshot of the system conditions. That means, the observed inflow into the DMA and the 12 pressures were available and used for evaluating the quality of calibration or leak detection and localisation.

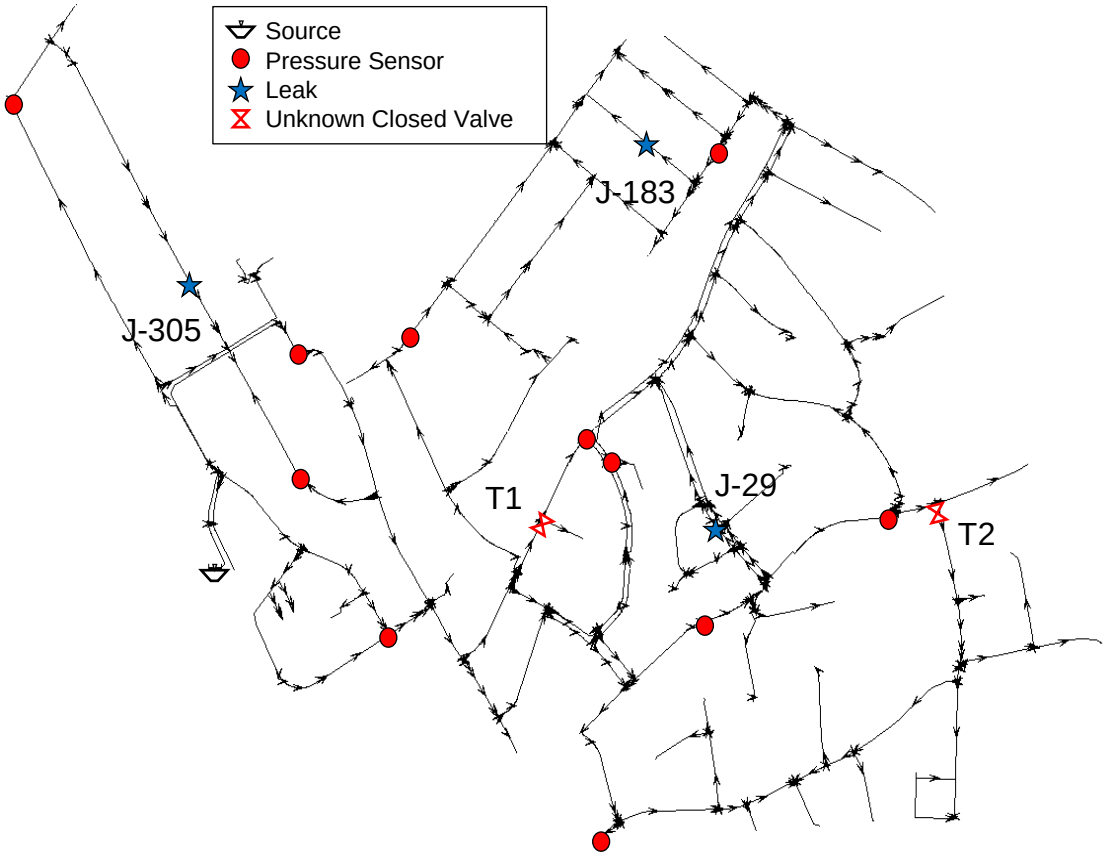

Figure 5.5. The SR2 Case District Water System.

5.2.2.3 Calibration and Leak Localisation Problem Set Up

The status of two randomly chosen valves (T1 and T2) was set incorrectly as "closed", to emulate a situation of unknown closed valves. This often happens in reality when the hydraulic model has not been updated with the true status of valves. This is considered as the actual system state without introducing any leakages for generating the artificial observations for model calibration. Based on this WDN state, all the datasets corresponding to each DF scenario were generated for subsequent calibration. Three leaks of different size were introduced at different locations to assess the quality of the resulting model after calibration based on a different DF and CAL scenarios. The locations are selected at nodes J-29, J-183 and J-305, as shown by the blue star symbols in Figure 5.5. This was done for two reasons: (a) to simulate a situation, when multiple leaks happen, or exist in WDNs (although normally one leak happens at a time); and (b) to test whether the leak localisation method is able to find multiple leaks of different size.

Table 5.3. Overview of the tested desktop calibration experiments.

Scenarios	DF1	DF2	DF3	DF4	DF5
CAL A	Random day with 0% demand variability	Random day with ±2.5% demand variability	Random day with ±5% demand variability	Random day with ±7.5% demand variability	Random day with ±10% demand variability
CAL B	Random day and de-noised with 0% demand variability	Random day and de-noised with ±2.5% demand variability	Random day and de-noised with ±5% demand variability	Random day and de-noised with ±7.5% demand variability	Random day and de-noised with ±10% demand variability
CAL C	Average day calibration dataset with 0% demand variability	Average day calibration dataset with ±2.5% demand variability	Average day calibration dataset with ±5% demand variability	Average day calibration dataset with ±7.5% demand variability	Average day calibration dataset with ±10% demand variability
CAL D	De-noised average day calibration dataset with 0% demand variability	De-noised average day calibration dataset with ±2.5% demand variability	De-noised average day calibration dataset with ±5% demand variability	De-noised average day calibration dataset with ±2.5% demand variability	De-noised average day calibration dataset with ±10% demand variability

The leak size at J-29 was 0.50 l/s, at J-183 0.95 l/s and at J-305 1.45 l/s, respectively. This totalled a global leakage of 2.90 l/s, which corresponds to approximately 20% of the average inlet demand.

The inlet flow for each of the five DF scenarios is plotted in Figure 5.6. Based on the DF scenario the average demand variation at any time step of a day considering the five day dataset is ±0 l/s (DF1), ±0.19 l/s (DF2), ±0.38 l/s (DF3), ±0.54 l/s (DF4) and ±0.73 l/s (DF5). For the idealized DF1 scenario, the demand pattern during all five days is kept the same. In contrast, the rest of the DF scenarios simulate a more realistic situation where each day's demand pattern is unique and the morning peak demand can vary up to ±0.42 l/s (DF2), ±0.83 l/s (DF3), ±1.19 l/s (DF4) and ±1.65 l/s (DF5), respectively.

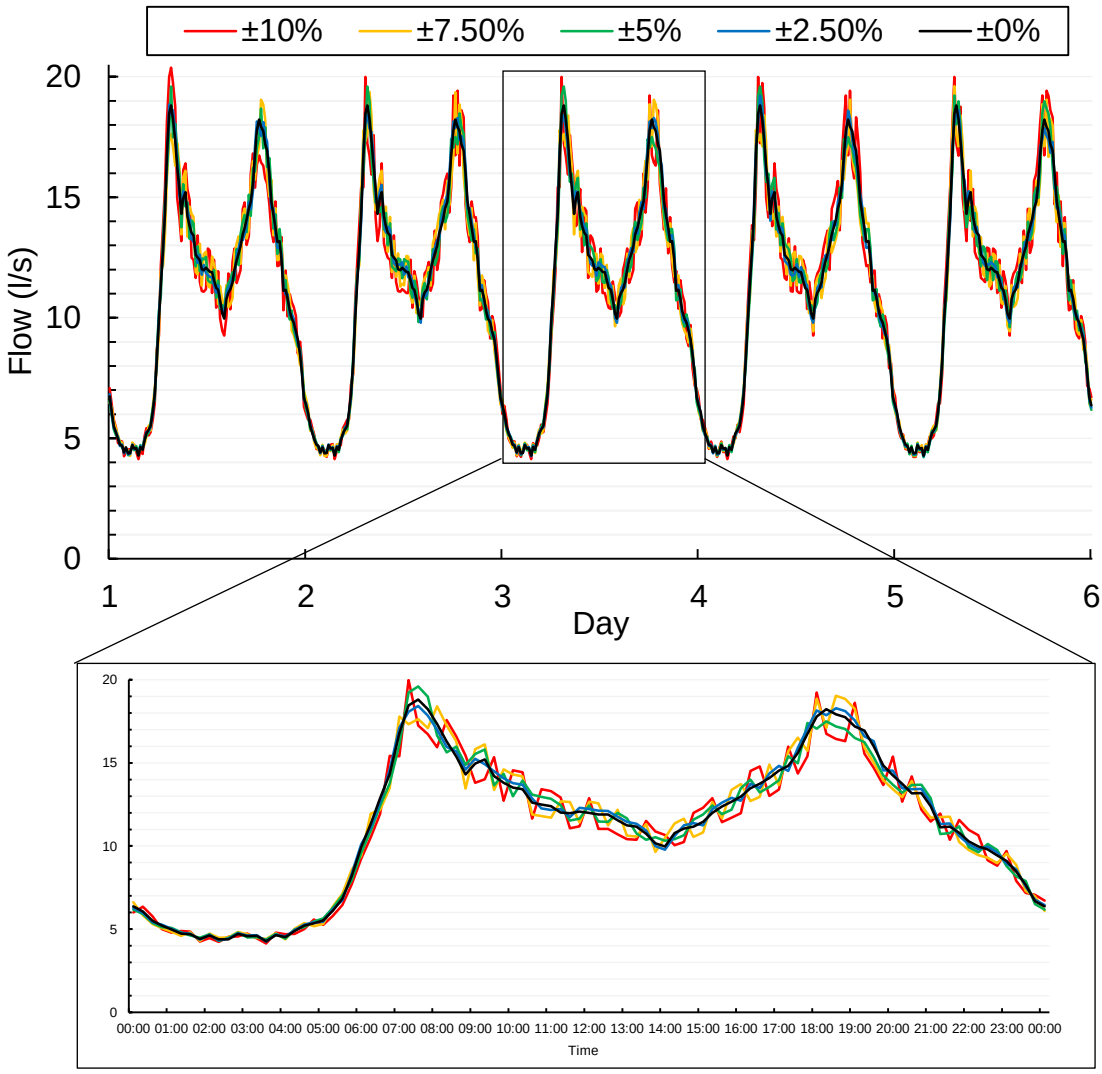

Figure 5.6. The generated inlet flow raw datasets for each DF scenario.

Four 24-hour calibration datasets were created for each DF scenario. For CAL A the third day of the dataset was randomly selected as the calibration dataset. For CAL B the third day as in CAL A was selected with the difference that the LOWESS de-noising technique was applied to the raw observations to before being used as the calibration dataset. CAL C used a calibration dataset after calculating the 24-hour raw data average over all five days. Finally, in CAL D the DPP stage was implemented, which created a 24 hour de-noised calibration dataset. An example of the four different 24-hour calibration datasets used in DF5 are illustrated in Figure 5.7. All generated datasets, including the leak localisation ones, involved a pressure error ranging between ±0.04m per recorded time step, for each device, and flow error within ±0.28l/s for the inlet pipe measurements.

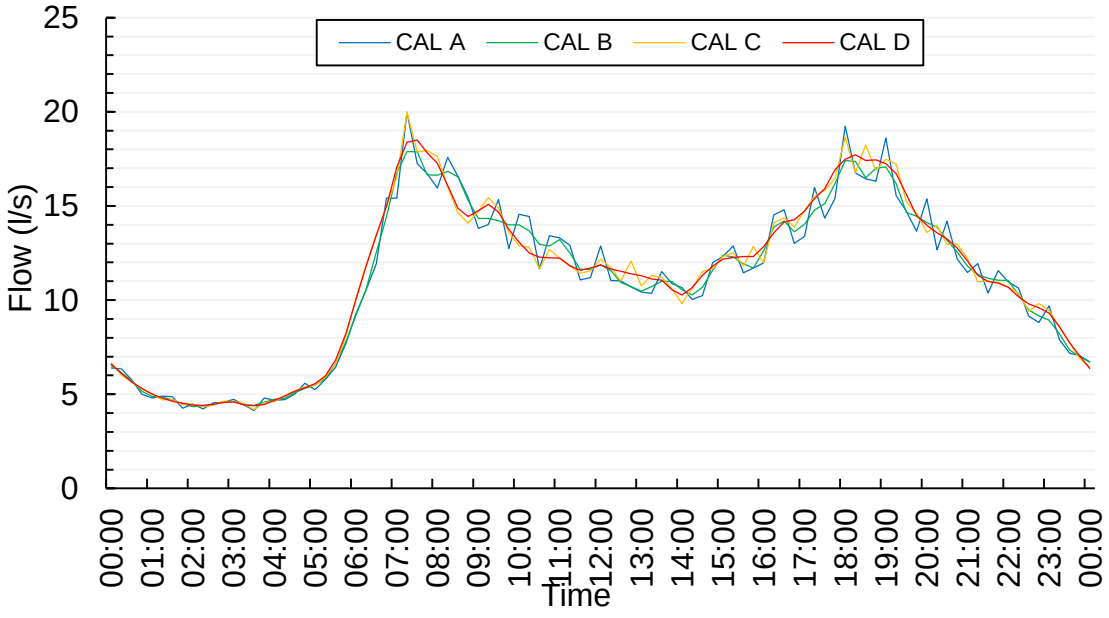

Figure 5.7. The calibration dataset for scenario DF5.

5.2.2.4 Calibration Performance

In the starting hydraulic model of each calibration process the status of valves T1 and T2 was set as “open”. In addition, the roughness values were artificially increased for 50 and 27 pipes associated with the CI and AC pipe material groups, respectively. In other words, their roughness value was increased to the next higher value from the available list of discretized roughness coefficients for each material. The location of the selected pipes was based on the fact that they occur on key flow routes where the average pipe velocity exceeds the average velocity across the entire system. In addition, the specific materials were chosen

considering that historically non-plastic and non-lined pipes have caused the majority of the model calibration problems. This is as due to the fact that their pipe roughness coefficient changes significantly faster than the lined and plastic pipes over their life cycle (Walski et al., 1988). Therefore, the roughness value of the selected pipes was adjusted to mimic the changes made to the roughness values during the traditional calibration procedure, by Modelling Engineers (Sage, 2018, Personal Communication). All CI pipes, except three that had a roughness value of 1.5mm assigned to them, had their values adjusted to 2mm. For the rest, the values were changed from 3mm to 3.5mm. All AC pipes had a roughness value of 0.03mm that was adjusted to 0.06mm. This was assumed to be the correct system configuration before implementing the calibration procedure for each DF and CAL scenario.

The water consumption profile in the hydraulic model is represented by a pattern of 96 15-minute coefficients. During the Demand Profile Calibration (DPC) part of the Profile Calibration (PC) stage the multiplier for the domestic consumption pattern was calibrated for all tested scenarios. In all DF and CAL scenarios the simulated inflow after DPC matched the corresponding calibration observations within the ±0.5% error over all time steps during the 24 hour period. As the system's pressure is not managed by any PRVs or pumps, the pressure profile is controlled solely by the DPC outcome. Consequently, there was no need for a Pressure Profile Calibration (PPC) part. Overall the PC stage took 2.5 hours to complete for each of the 20 scenarios. The resulting hydraulic model was then used as a baseline for the Search Space Reduction (SSR) stage.

The SSR stage considered all valve and pipe components of the WDN hydraulic model (sections 3.3.3-3.5.3 and 3.3.4-3.5.4). Following Step 1 of SSR (section 3.3.3), the simplified problem was reduced to 74% the original valve components, i.e., 131 out of 178 valves. Then, at Step 2 (section 3.4.3) the search space was further restricted to 37% of components, or 65 detectable valves. Step 2 required about five minutes to complete. Following Search Space Optimisation (SSO) at Step 3 (section 3.5.3), the fittest scenario in all cases indicated the existence of a maximum two closed valves in the WDN. The process took about 20 minutes to run for each scenario after restricting the maximum number of possible valves

to $v = 10$ as a means to expedite the analysis. The grouping procedure at Step 1 (section 3.3.4) resulted in a total 21 different pipe groups. The CI, AC and HPPE material groups were divided into 11, 3 and 7 diameter groups, respectively (Table 5.4). Following Step 2 (section 3.4.4), only four pipe groups from the CI set remained in the search space after determining the Detectable Pipe Roughness Coefficient groups. Step 2 for the pipe components was completed within five minutes. In all scenarios the pipe groups 1, 3, 7 and 8, remained in the search space, which constitute 215 out of 433 pipes in the WDN (Table 5.4). This resulted in a reduction of more than 50%.

Table 5.4. The Candidate Grouping after Step 1 of SSR Stage.

Group Number	Material	Diameter (mm)	Components	Pipe Group Length (m)
1	CI	66	110	3763
2	CI	76	4	23
3	CI	92	68	1528
4	CI	99	1	2
5	CI	101	18	252
6	CI	106	2	61
7	CI	117	5	653
8	CI	142	32	716
9	CI	152	2	4
10	CI	157	5	104
11	CI	209	3	69
12	AC	75	4	69
13	AC	97	79	3888
14	AC	225	4	57
15	HPPE	50	17	343
16	HPPE	73	19	686
17	HPPE	101	30	885
18	HPPE	146	11	402
19	HPPE	200	1	9
20	HPPE	203	14	633
21	HPPE	233	4	451

During the Component Calibration (CC) stage, two unknown closed valves were detected in all scenarios. However, none of the scenarios detected T1 and T2 as closed. Instead, the optimum result in each scenario identified the two valves just downstream of actual closed valves, T1 and T2, i.e., at a distance of 110m and 50m, respectively. Although the true valve locations were not detected, the flow paths on each closed valve is located were found correctly. This allows for an inspection of the pipeline at a street level and identification of the actual closed valves. On the other hand, compared with the actual system state, the pipe roughness calibration outcome led to a smoother roughness coefficient (0.88mm or 1mm) for the smaller diameter groups (i.e., 66mm and 92mm). Conversely, a rougher coefficient (3.5mm, 5mm or 7.5mm) was selected for the larger diameter group (i.e., 116mm and 142mm). For all 20 tested scenarios, the calibration criteria (section 2.4.5) for all error bands were met for 100% of simulated outputs. Therefore, for all CAL and DF scenarios the resulting calibrated model matched the corresponding measurements within ±0.5m. Overall the CC stage required one hour to complete for each scenario.

5.2.2.5 Leak Localisation Accuracy

The LIM was, then, implemented for each CAL scenario and was tested on how well in performed for detecting and localising the three leaks for every DF scenario. The simplification step reduced the search space in all scenarios to 40% of network nodes, i.e., to 233 out of 589 nodes. Following Part I of SSO, a leak of size 2.90 l/s was detected in each case. Any node where the MDNL is higher than the detected water losses was eliminated from consideration. This reduced the search space further to 38% of nodes as potential leak locations. The average MDNL for the 225 candidates was approximately 0.48 l/s in all 20 desktop experiments (Table 5.3) where the model was calibrated before leak detection and localisation. This corresponded to six different scenarios for investigating the maximum possible number of leaks in the WDN, which ranged from a single large leak of 2.90 l/s up to six smaller leaks of 0.48 l/s. In all 20 cases Part II optimisation analysis showed that the leak scenario with the minimum objective function value was the existence of either three or four leaks in the WDN. The SSR stage took a total time of about one hour to complete. LDL was, then, run to find the size and locations of a maximum of three or four leaks.

This depended on the outcome of Part II for each tested scenario. The LDL stage was completed after two hours of hydraulic simulation evaluations.

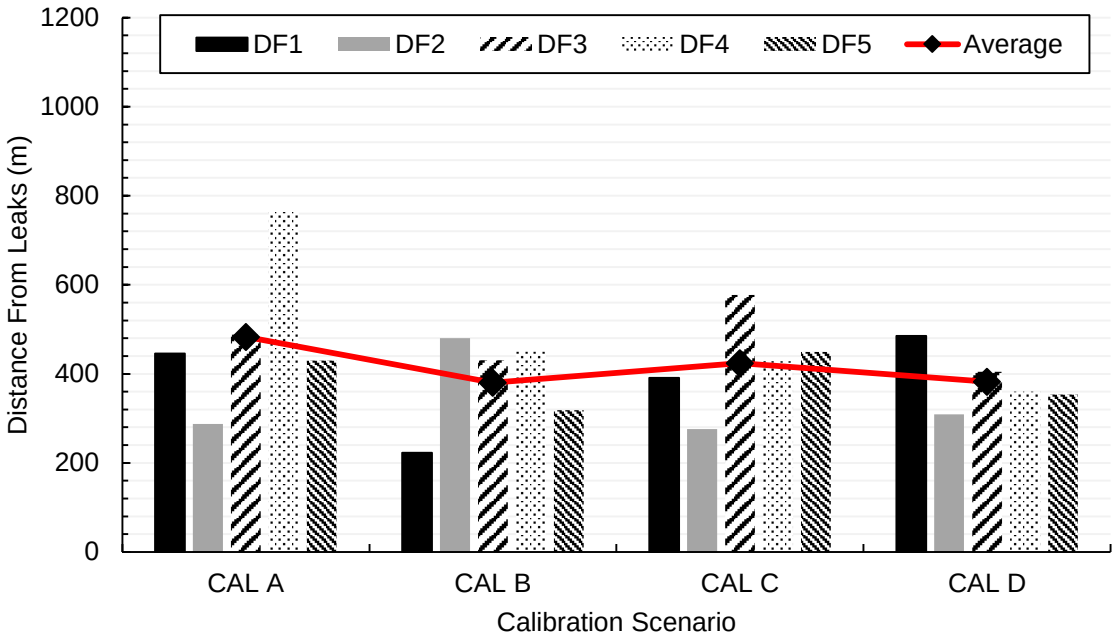

Figure 5.8. Average distance between all the detected leaks compared to the actual leak locations, for each calibration and demand fluctuation scenario.

Figure 5.8 compares the average distance between a detected and an actual leak location in the WDN for each CAL and DF scenario, averaged from the results of all three leaks. In general, in CAL B and D cases, where the model was calibrated using a de-noised dataset, the leak localisation performed better relative to CAL A and C. On average, for CAL B and D cases, a leak is reported between 60-100m closer to the actual leak location. CAL A, which assumed a dataset similar to the currently used across the UK Water Industry can only achieve a good leak localisation when demand fluctuation in the DMA is low, i.e., at DF2. On the other hand, CAL B produced the best results for the ideal scenario DF1 and the highly variable scenario DF5. It appears that the de-noising technique reduced the noise caused by the measurement error and demand fluctuation in DF1 and DF5, respectively, which helps the technique to find the leak locations. The test case CAL C achieved the best leak localisation accuracy relative to the rest of scenarios at DF2. This is because the noise in measurements and demand was reduced and, thus, producing less impact on the leak localisation outcome. Finally, CAL D achieved the best results at DF3 and DF4, while similarly good results were also achieved at the DF2 and DF5 demand fluctuation scenarios.

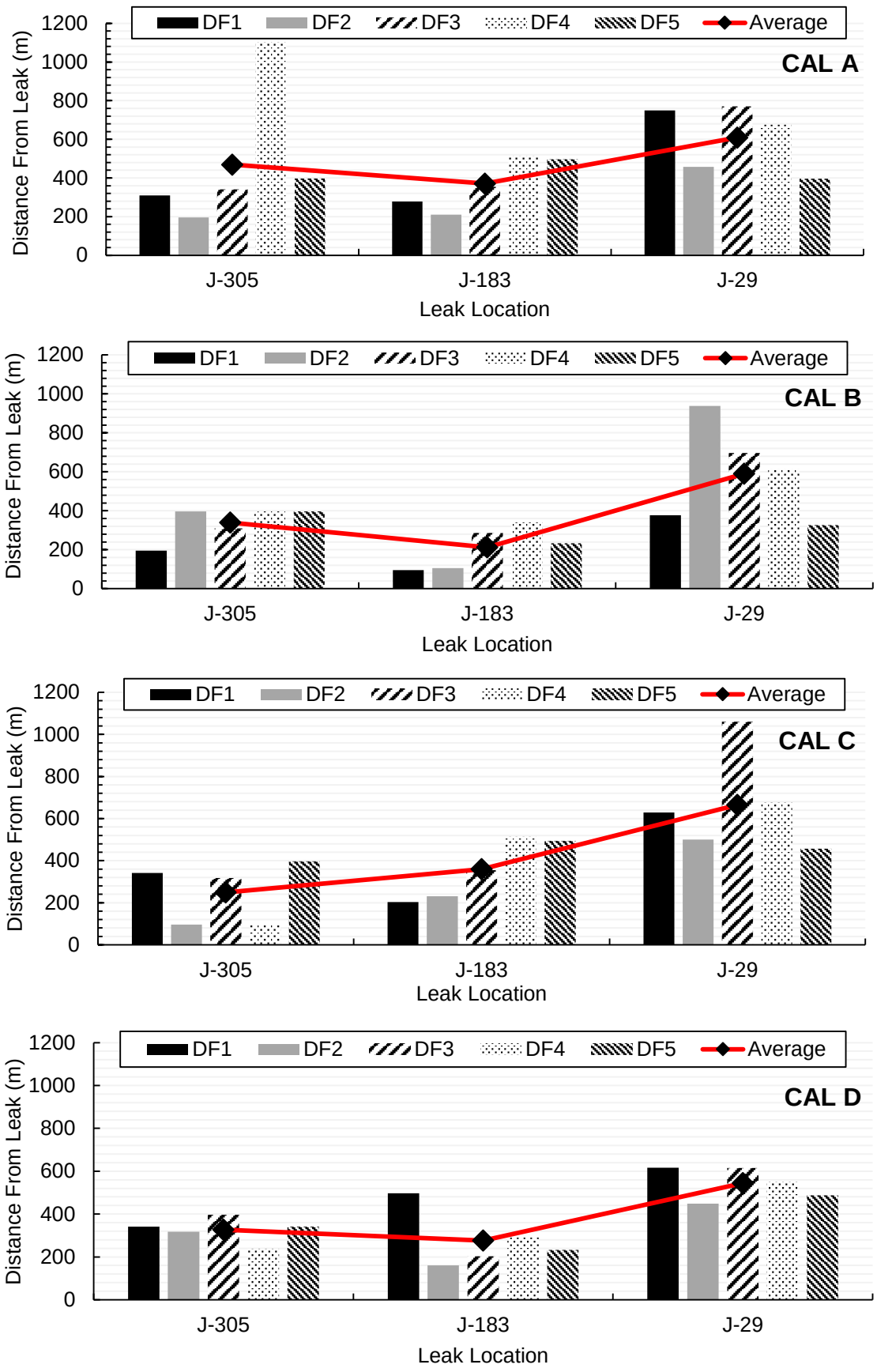

Figure 5.9. Distance between each detected leak location relative to the actual leak locations, for each calibration and demand fluctuation scenario.

A more detailed comparison on the leak localisation results for each reported leak is presented in Figure 5.9. The distance from each detected leak is compared with the actual leak locations J-305, J-183 and J-29 for each CAL scenario and with respect to the DF scenarios. CAL B and D performed equally well in finding the largest leak J-305. The average distance from the actual leak location across all DF scenarios was 339m (σ = 89m) and 327m (σ = 58m), respectively. However, CAL C achieved the best localisation for J-305 at an average distance of 250m (σ = 143m), while CAL A had the worst performance with 470m (σ = 362m) error. CAL B and D achieved the best results in localising J-183 within 212m (σ = 109m) and 277m (σ = 132m), respectively, as opposed to the 360m (σ = 144m) and 370m (σ = 133m) distance as a result of CAL A and C. The leak localisation for the smallest leak J-29 failed in all 20 desktop experiments, while, CAL D produced the best results with a 545m error (σ = 75m). A leak was detected in all cases. However, the similar size of the leak J-29 with the average MDNL value, i.e., 0.48 l/s, combined with the noise in measurements and/or the demand fluctuation severely affected its localisation. Overall, CAL D demonstrated the most robust results for all three leaks with the least variability in the reported leak distance, when demand fluctuation is larger than 0%, i.e., for all realistic scenarios.

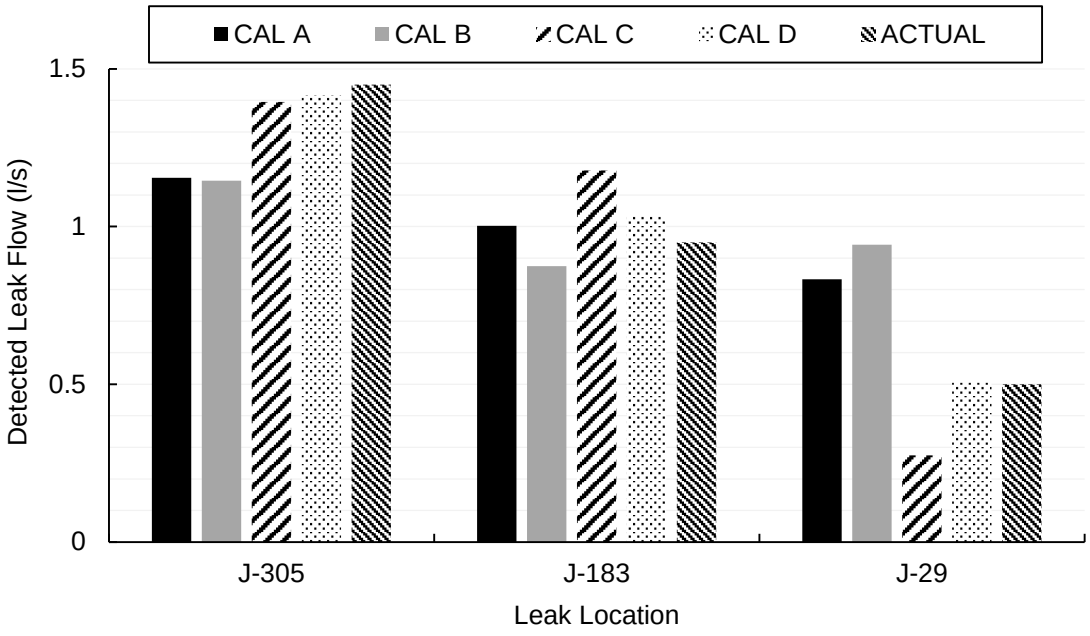

Figure 5.10. Comparison of reported leak size for each CAL scenario, averaged for scenarios DF2 – DF5.

The robustness of the CAL D case is also supported by the comparison in the detected leak size averaged for all DF scenarios, as presented in Figure 5.10. Although in all scenarios the global leakage value was correctly identified as 2.90 l/s, the size of each reported leak differed for every CAL scenario. Overall, the CAL D case resulted in a leak size closer to the actual leakage values in all realistic demand variability scenarios, i.e. for DF2-DF5. The size of leak J-305 could be more accurately detected following CAL D (1.42 l/s) and C (1.40 l/s), while the size of leak J-183 following CAL A (1.00 l/s) and D (1.03 l/s). Finally, the size of leak J-29 was simulated more accurately by CAL D (0.51 l/s).

Interestingly, based on the optimum solution obtained for each tested scenario, in all desktop experiments CAL C led to a better match of the model outputs with the leak localisation data compared to all other CAL scenarios. This is presented in Figure 5.11, which compares the normalized objective function values after leak localisation for each CAL and DF scenario. This comes to an opposition with the actual results and raises a possibility of overfitting. It can be observed that for the CAL B and D cases, which used a de-noising technique for model calibration, that they do not fit the observations well in the idealized scenario DF1. Furthermore, their fitness increases at a higher demand uncertainty. This is because when demand fluctuations are introduced (i.e., random error), the model that is calibrated based on smoothed data can more accurately simulate the general consumption pattern, as fluctuations are eliminated.

On the other hand, at a “zero-fluctuation” situation when the demand pattern is reliable, smoothing the data causes more uncertainty, as noise is being introduced instead of being removed. Finally, calibration based on CAL A, which mimics the current practice across the UK water industry, can lead to a good match with the leak observations only in an ideal demand fluctuation scenario, i.e., DF1. On the other hand, in a more realistic situation with larger variability, CAL A produces the worst fit with the leak observations, compared to the rest of the calibration scenarios. This can be a result of the bias introduced to the calibrated model after subjectively choosing a calibration dataset of a unique day from the raw observations.

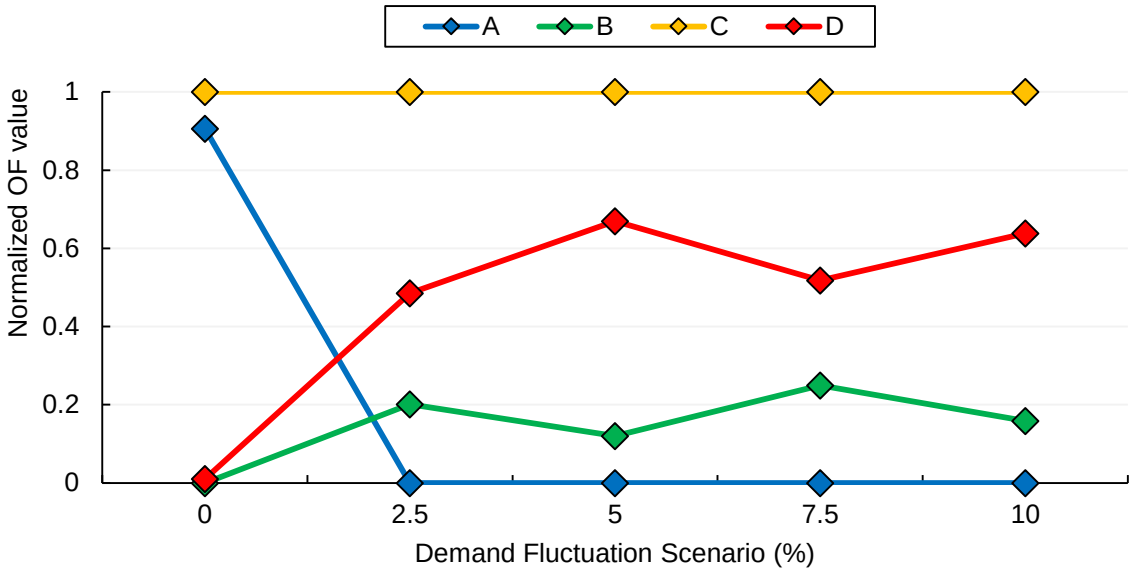

Figure 5.11. Comparison of the normalized objective function value (0-worst, 1-best) achieved after leak localisation for each CAL scenario as a result of changing demand fluctuation.

5.3 Real District Water System Applications

5.3.1 Case Study R1

5.3.1.1 System Overview

The LIM was tested further on a real UK DMA (Figure 5.12), which had historically high levels of leakage recorded. This case also used data of a real leak event, recorded by pressure and flow sensors deployed by the utility before the event happened. The system is fed by a single source that serves the domestic consumption associated with a rural area of approximately 1,000 properties. Flow from the source node varies normally between 2.6 l/s at MNF and 8 l/s during the morning peak demand. The hydraulic model contains 601 nodes, 461 pipes and 159 valves (not presented in Figure 5.12). The length of the network is 45km. Out of the total number of pipes, 55% (254 pipes) are Asbestos Cement (AC), 9% (42 pipes) Cast Iron (CI) and 36% (165 pipes) Polyethylene (e.g. PE, MDPE, HPPE). The DMA has flow and pressure sensors at the inlet, and 10 inner pressure sensors. Typically, the utility deploys one pressure logger per 200 properties. However, with the aim to introduce real time modelling of WDNs, the system's logging density was increased to roughly one logger per 100 properties. The DMA contains three Pressure Reducing Valves (PRVs) due to the large range in

elevation between the lowest point at 28m and the highest at 221m. The result of the elevation differences is a huge pressure range of 20 to 148 metres of head (average of 53.8m).

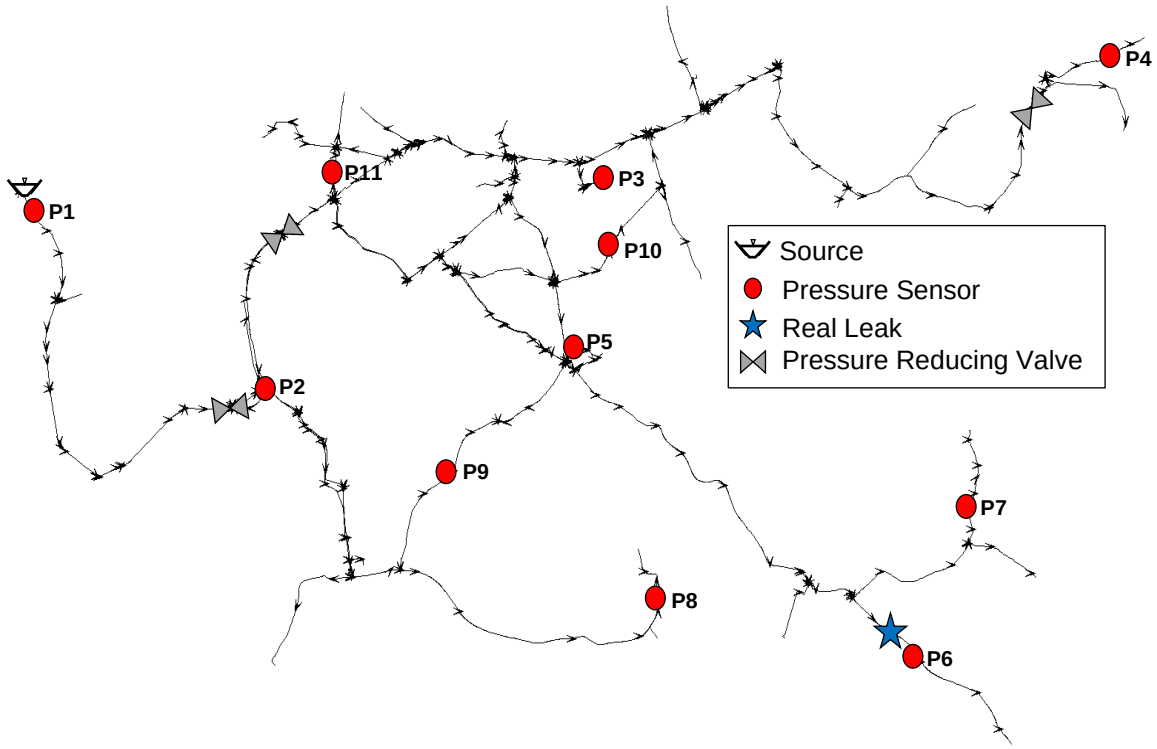

Figure 5.12. The R1 Case District Water System.

5.3.1.2 Available Data

The “raw” dataset of observations comprised of pressure and flow readings that were collected for 17 days for the period between October 31 and November 16, 2016. The LIM process used data from a historical leak, which has been detected and localised by the utility. The leak started at 15:30 hours on November 7, 2016, and lasted for around seven days. The exact location where it was found and repaired by the leak detection engineers is indicated by the blue star in Figure 5.12. A systematic flow difference in the range of 2.15 and 2.5 l/s was observed at the MNF throughout the seven days when leak was running, where the inlet flow increased from 2.7 l/s to 5.2 l/s. Moreover, the leak caused a significant head loss around the leak location of more than 30m (Figure 5.13). Interestingly, a large variation was observed during the morning peak hours, as the leak impact was not as obvious, compared to the night flow increase. The leak effect only caused an increase of approximately 1 l/s compared to the day before the leak

started. This demonstrates the large uncertainty introduced in the leak detection and localisation process when the majority of the observed flow comes from customer consumption and not leakage.

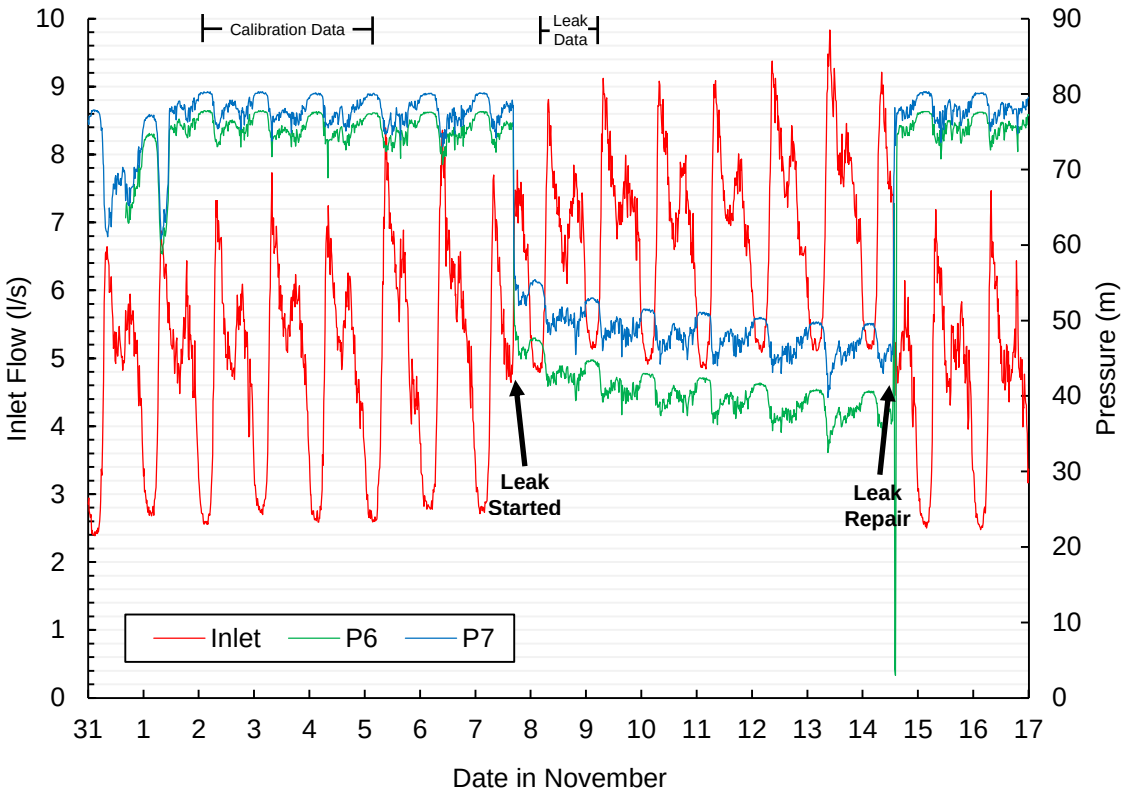

Figure 5.13. The time series field record of the inlet flow meter and pressure data from the closest sensors to the leak location.

To assess the leak localisation methodology, a calibrated average day model of the system according to the company's standards was necessary. In other words, a model that simulated pressures at the sensor locations with an accuracy range of ±2m relative to the observations. However, this was not available to the water company at the time of the leak. Therefore, the collected measurements before the leak occurred were used to calibrate the model using the CM so it can be used for finding the leak. The observations for the calibration part involved the time series of flows into the system and the pressures at 11 locations, between November 2 and 4. The collected measurements before the November 2 were not taken into account due to an unstable demand and pressure profile. This was as a result of an operational work in the WDN that had ended in the afternoon of

November 1. In addition, the collected data on November 5 and 6 involved a different consumption pattern due to the weekend, thus, they were not considered for simulating the average hydraulic conditions. A total of 96 calibration measurements over a 24-hour period were used for the calibration of the average day hydraulic model.

On the other hand, the observed data for the leak detection and localisation part involved the time series of flows and pressures on November 8, i.e. one day following the leak start. This was to test whether the methodology could detect the leak location and flow at an earlier stage, compared to when it was really discovered and repaired, on November 14. A total of 96 field observation data sets over 24 hour period (from midnight of November 8 to midnight of November 9), have been imported into the optimisation modelling tool of the LIS.

5.3.1.3 Calibration Method Implementation

5.3.1.3.1 Data Pre-Processing

The system demand for the three-day dataset of raw measurements varies within ±4.6% at any time step with respect to the average demand. Therefore, it was decided that a calibration approach, which uses a de-noised dataset of the average day, i.e., similar to CAL D scenario (Table 5.3), would be used for the model calibration process. The results from Case Study SR2 suggest that a more accurate representation of the average day conditions can be produced due to the large demand fluctuation. This can lead to improved leak localisation accuracy, due to the reduction in noise caused by demand fluctuation combined with the measurement error. The DPP stage took about 15 minutes to complete. Following the DPP stage, the readings from the three selected days were converted to a 24-hour dataset of the average conditions for subsequent model calibration. The outcome of the DPP stage is illustrated in Figure 5.14. The high variability manifested during each day results in a noisy demand pattern. This is due to the abrupt changes in consumption at each time step of the 24 hr dataset of the average conditions. However, using the LOWESS smoothing technique the demand-related noise was removed. The outcome was a de-noised calibration dataset that was used within the CS.

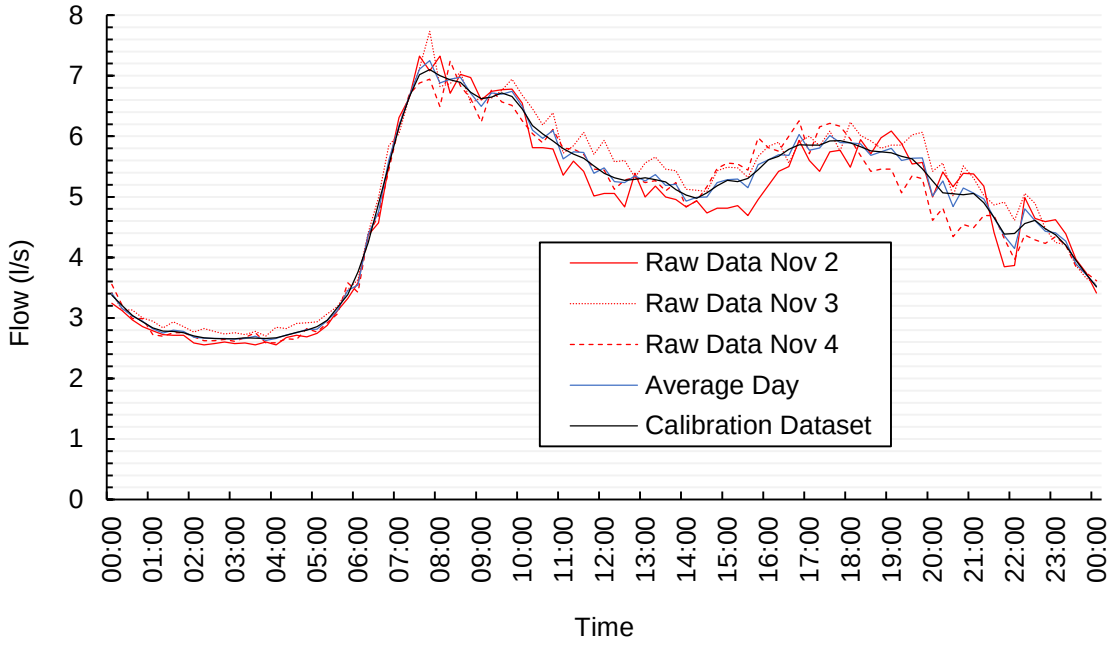

Figure 5.14. Comparison of the resulting average day dataset and the calibration dataset after the DPP stage, relative to the raw observations.

5.3.1.3.2 Profile Calibration

During the DPC part, the 96 multiplier coefficients that simulate the domestic consumption pattern were calibrated. Although a similar pattern exists for simulating background leakages in the WDN, that pattern was not included in the calibration procedure.

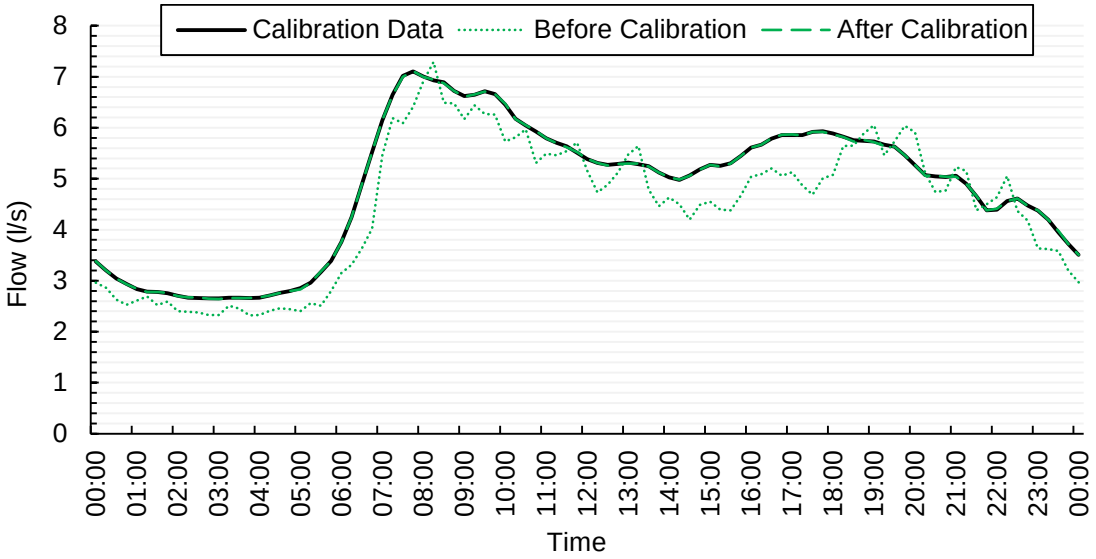

Figure 5.15. Comparison of the updated hydraulic model before and after Demand Profile Calibration as well as relative to the calibration dataset.

This was decided assuming that there was no change in the amount of background leakage, with respect to the day before the leak happened. The DPC part was completed within three hours. The simulated inflow after DPC matched the calibration observations within ±1% error (typical flow meter error) over all 24 time steps (Figure 5.15). An improved representation of the average hydraulic conditions during the same period was also achieved.

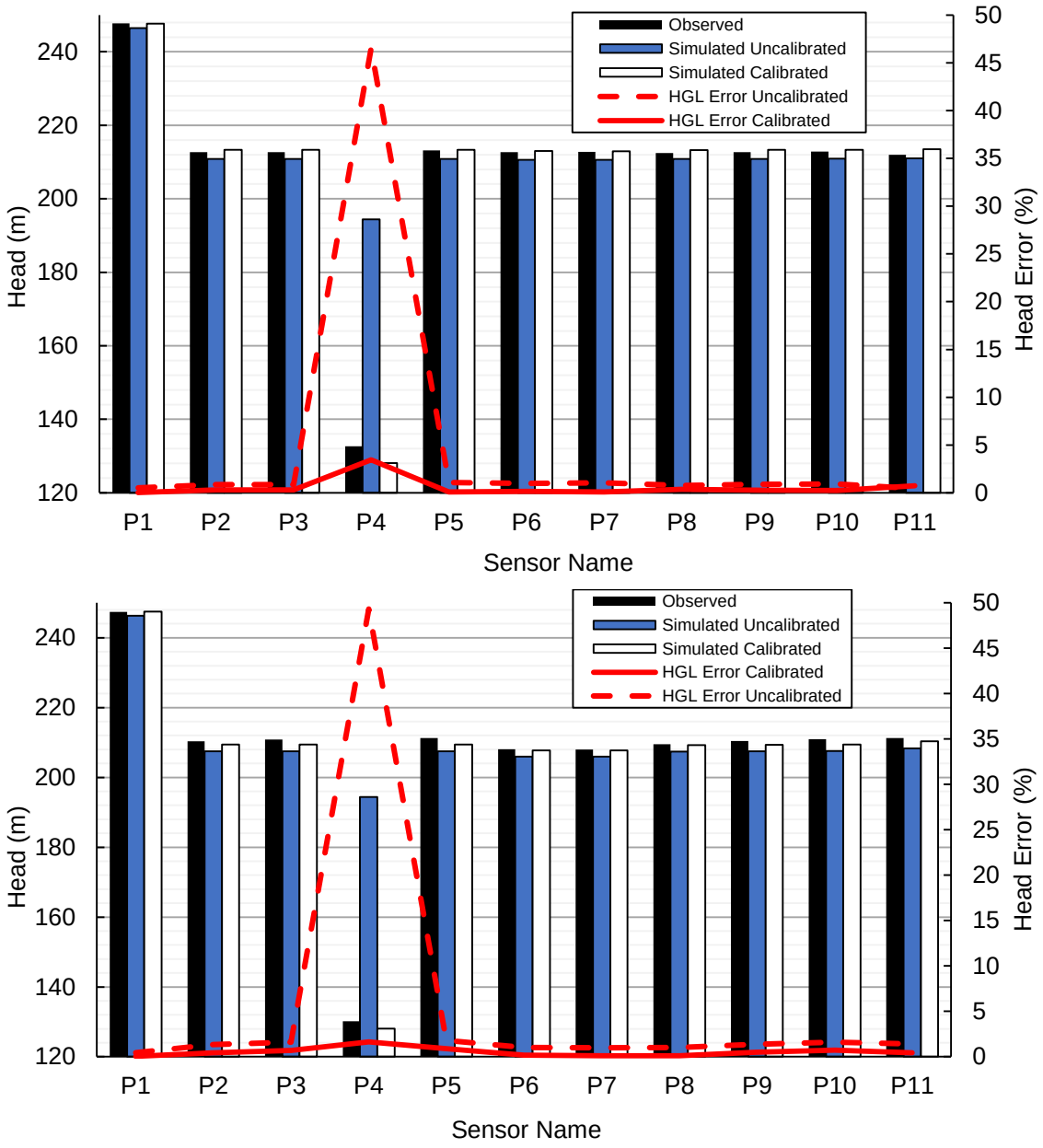

Figure 5.16. Comparison of the observed and simulated heads for the uncalibrated model and optimised calibration solution during MNF at hour 3 (top) and the morning peak at hour 8 (bottom).

Then, at PPC, the setting of the three PRVs was calibrated for each time step. The PPC part also took approximately three hours to complete. This ultimately led to a calibrated hydraulic model at a macro-level. In other words, observations are close to, but do not meet the desired performance criteria of ±2m accuracy relative to the observations. A comparison of the simulated heads at the sensor locations before and after calibration is given in Figure 5.16. The figure also provides the observation during the hours of MNF and the morning peak, i.e. at 03:00 and 08:00 hours, respectively. The error achieved after PPC is less than 1% across all sensors during both the MNF and the morning peak. The exception is P4 at MNF, where a difference of around 3% exists between the observed and simulated head values. Although this does not represent a fully calibrated model, it serves as a very good baseline for further calibration during the Component Calibration stage.

5.3.1.3.3 Component Calibration

Simplification reduced the search space to 57% of network valves, i.e. 90 out of 159 valve components. The sensitivity analysis of valve status changes further restricted the candidates to 27% of the WDN, or 43 Detectable Valve Locations. The analysis was completed within five minutes. Prior to any optimisation run for identifying the detectable valve locations, the system evaluation was conducted by comparing the field observed heads with the simulated heads over 24 hours. This provided a starting point for defining the boundary conditions error, i.e., the objective function value when no valve status change is made in the WDN model. Considering the large inconsistency in demand during the hours when the majority of demand comes from customer consumption, an increased weight was given to the pressure and flow differences during the hours of low demand. This involved hours between 00:00 and 06:00 and was applied to all optimisation analyses that were performed. The weight value for the pressure and flow differences at each time step was derived by taking into account the standard deviation for each set of measurements during the considered hours, compared to the 24 hour standard deviation, i.e., for the whole day. The closer the standard deviation between 00:00-06:00 to the 24 hour value, the lower the weight given to the differences. In order to expedite the search space reduction process, a user specified value was used for the maximum number of closed valve scenarios

based on the expert knowledge of the system flow paths. In this case, the maximum number of possible closed valves was defined as $v = 10$. Search Space Optimisation showed that the most likely valve scenario was the existence of one closed valve in the WDN. The analysis was carried out within 30 minutes. On the other hand, the grouping procedure resulted into 19 groups of pipes. They were grouped based on the material and diameter, as no age information was available for the pipes of the specific DMA. The possible roughness coefficients for each group were established from the company's records depending on the material (Table 5.5). From those, only two of the pipe groups (2 and 11) remained in the search space after determining the Detectable Pipe Roughness Coefficient Groups. The analysis for the Detectable Pipe Roughness Coefficient Groups took about five minutes. Therefore, out of the total 461 pipes in the WDN model, only the 211 that constitute the two remaining groups were considered for calibration. This is because any change in the roughness value of the rest of the groups did not produce head loss larger than the sensor accuracy range.

Table 5.5. The Candidate Grouping after Step 1 of SSR Stage.

Group Number	Material	Diameter	Components	Pipe Group Length (m)	Candidate Roughness Values (mm)
1	AC	50	5	543	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
2	AC	75	169	21552	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
3	AC	76	41	1468	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
4	AC	97	11	527	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
5	AC	101	11	619	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
6	AC	122	6	475	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
7	AC	146	2	4	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
8	AC	150	1	11	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
9	AC	152	3	4	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
10	AC	157	5	14	0.3, 0.2, 0.1, 0.08, 0.06, 0.03
11	CI	209	42	3784	7.5, 6, 5, 4, 3.5, 3, 2.5, 2, 1.5, 1, 0.88, 0.03
12	HPPE	50	17	2187	0.01
13	HPPE	73	47	3730	0.01
14	HPPE	79	7	1891	0.01
15	HPPE	81	5	58	0.01
16	HPPE	101	1	3	0.01
17	HPPE	106	7	7	0.01
18	HPPE	158	41	4944	0.01

19	HPPE	207	40	3119	0.01

To verify this, the diameter pipe groups were reconstructed by splitting the AC pipes into three groups based on user specified diameter ranges. In addition, all CI pipes and HPPE pipes were considered as additional two groups. This reduced the initial 19 pipe groups to five pipe groups before rerunning Step 2. Group 1 comprised of AC pipes with a 'small' diameter ranging from 50-80mm, Group 2 of AC pipes with a 'medium' diameter between 81-130mm, and Group 3 of AC pipes with a 'large' diameter of 130mm and above. Group 4 comprised of CI pipes and finally, Group 5 of HPPE pipes. Again, only Group 1 and 4 remained in the search space after the pipe group reconstruction. Compared to the change in pressure caused individually by pipe group 2 and 11 (Table 5.5) there was no significant difference when Group 1 and 4 were tested. This indicated that the main impact on pressure is caused by pipe groups 2 and 11. The SSO at Step 3 (completed within ten minutes) indicated that all possible combinations with the remaining pipe groups were increasing the objective function value, i.e., worsening the fitness. Therefore, no pipe groups were considered in the Component Calibration stage. The additional decision variable was the location of a one possible closed valve.

Table 5.6 illustrates the optimisation outcome following the Component Calibration stage. One valve was found to be closed, which was verified in the field following the modelling analyses. Compared to the starting uncalibrated hydraulic model, where only 30% of simulated outputs met the desired criteria, the updated model matched the pressure observations much better over 24 hours. The desired company's criteria were met for the error band of ±2m, i.e. for 100% of simulated outputs. On the other hand, the error bands of ±0.5m and ±0.75m associated with the criteria in Section 2.4.5., were missed, but only for few cases.

Table 5.6. Simulated Pressures relative to the calibration accuracy criteria.

Error Range (m)	±0.5	±0.75	±1	±2
Before Calibration (%)	0	0	1	30
After Calibration (%)	83	94	98	100

5.3.1.4 Leak Inspection Method Implementation

5.3.1.4.1 Leak Search Area Reduction

The simplification methodology reduced the search space to 47% of network nodes. Prior to any optimisation run for identifying possible leakage hotspots, the system evaluation was conducted by comparing the field observed inflows with the simulated inflows of actual consumption over 24 hours. This provided a starting point for defining the MDNL flows for each candidate leak location. The MDNL was computed based on a typical reading error of ±0.1% associated with the utility's devices. The MDNL analysis was completed within ten minutes. Considering the large variation and inconsistency in demand during the morning peak hours and throughout the day, an increased weight was given to the pressure and flow differences during the hours of low demand, i.e., between 00:00 and 06:00. Following Part I of the optimisation-based search reduction, a leak of size 2.25 l/s was detected. This corresponds to around 33% of the average inflow readings. The leak size in Part I was detected after ten minutes of calculations. After eliminating the nodes with MDNL values higher than the detected leaks, the average MDNL was approximately 0.13 l/s. This corresponded to 18 different scenarios for the possible number of leaks. This ranged from a single leak of 2.25 l/s up to 18 smaller leaks of 0.125 l/s.

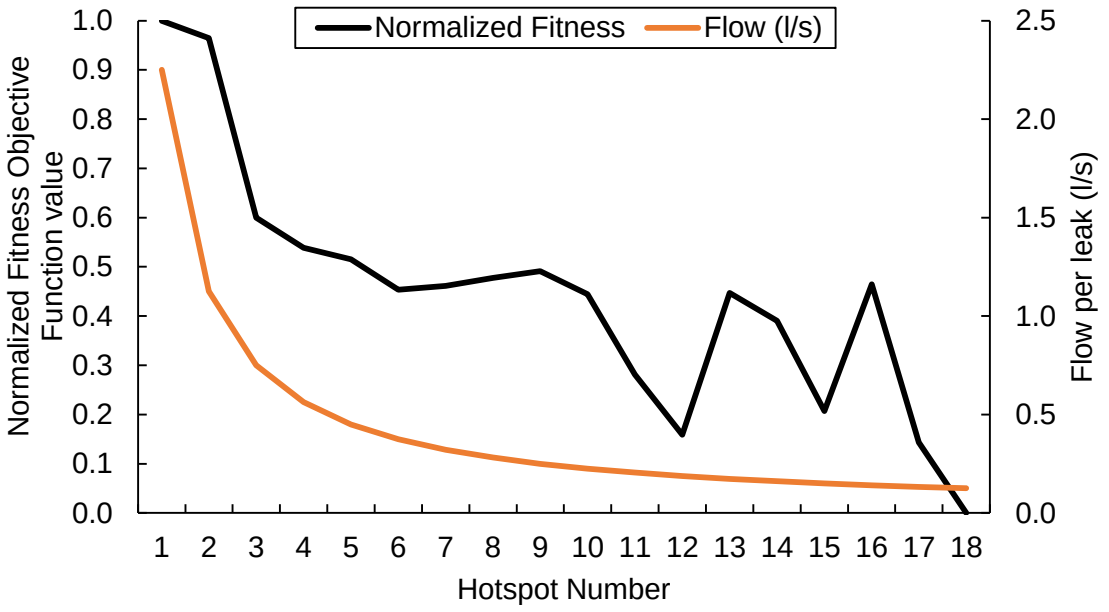

Figure 5.17. Part II optimisation analysis outcome for the different leak scenarios in the WDN (0-worst, 1-best).

Part II optimisation analysis was completed within 45 minutes and showed that the most likely leak scenario was the existence of a single large leak in the WDN. This was the scenario with the best objective function value (Equation 3.16 – Section 3.5.2), as shown in Figure 5.17. By removing all nodes where the MDNL value was larger than 2.25 l/s, the final list of leak location candidates resulted in 27% of the network nodes. LDL was, then, run for finding a single leak, while the setting of the three PRVs was allowed to vary by ±10% to emulate the response of a PRV setting to a real leak. The analysis took about two hours to complete.

The optimum solution indicated a node on the branch where the real leak occurred, but 800m downstream from it (Figure 5.18). Considering the length of the WDN mains this is an error of 2% (by length). This suggests that the approach was able to correctly identify the leakage area within a small distance of a true leak. Furthermore, the search for finding a leak was narrowed down between sensors P6 and P7, which corresponds to a maximum search distance of only 10% of mains length. As the presented methodology used a 24-hour sensor data to detect the leak and that in reality it took seven days for the utility staff to locate the leak, the new approach can possibly contribute to a much earlier detection and localisation of the leak.

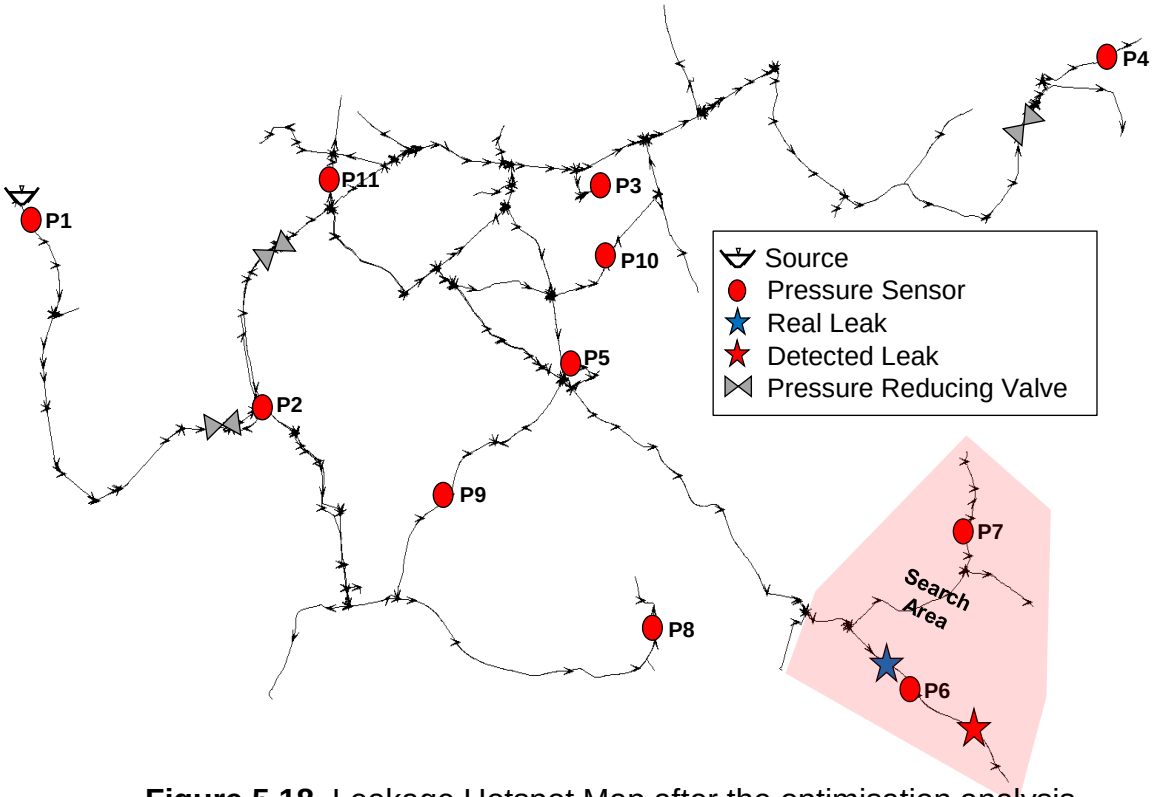

Figure 5.18. Leakage Hotspot Map after the optimisation analysis.

5.3.1.4.2 Model Calibration

The fact that the detected leak in Figure 5.18 was reported downstream, instead of upstream of the sensor P6, could exhibit errors in the calibration process. It is possible that the roughness value of all the pipes between sensors P5 and P6 that belong to Pipe Group 2 (Table 5.5) considered for calibration, were set at a higher value than it should be. This was done to compensate for the head loss and match the observed pressures within ±1m, before the leak happened. The results suggest that additional measurements upstream of P6 are necessary to get more accurate value for the pipe states. Nevertheless, an EPS analysis was completed with the updated hydraulic model, i.e., after simulating the identified leak. The simulated inflow after LDL matched observations much better than the original model over 24 hours (Figure 5.19). This is evident especially between 00:00 and 06:00 across all days where an increased weight (approximately x10 times larger) was assigned to pressure and flow differences. The leak localisation dataset involved measurements on the 8th of November. Interestingly, by comparing the updated inflow data in Figure 5.19 with the inflows between November 9 and November 14, before the leak was found and repaired, an even better match is produced. However, the hydraulic model outputs do not match with the weekend demand on November 12 and 13, as it follows a different profile, relative to weekdays.

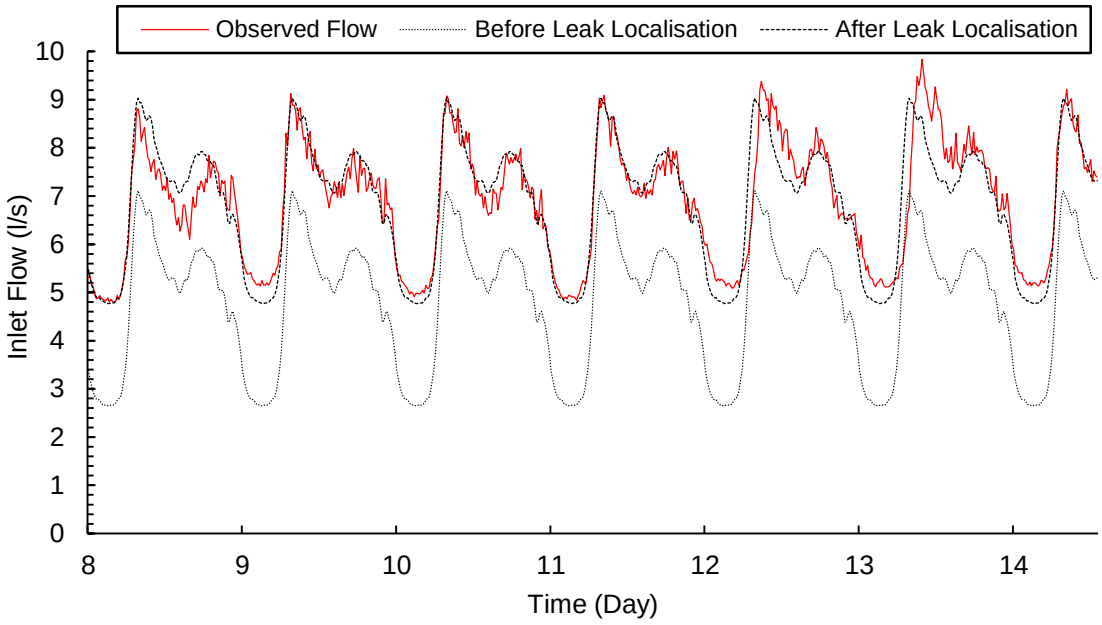

Figure 5.19. The Flow Differences before and after leakage detection.

A comparison of the observed and simulated heads at the sensor locations during the hours of MNF, at 03:00 hours is also given in Figure 5.20. From the calibration point of view, it can be observed from this figure that the error is equal to or lower than 1% at most sensors. The exception are P6 and P7, where a difference of around 3% exists between the observed and simulated head values. Although this does not represent a fully calibrated model, it serves as a very good baseline for further calibration, following the localisation of the leak in the field. Moreover, the approach successfully narrowed down the leak search space and could contribute to earlier leak localisation. A point to raise considers the fact that PRVs follow a profile variation at their setting throughout the day. This functionality could not be replicated by the EPANET model used during the leak localisation. Thus, the profile of the response in the setting of the three PRV's could not be simulated. However, this does not invalidate the Leak Detection and Localisation methodology as the leak size and location could still be identified due to the large leak impact on pressure of at least 30m. On the other hand, if the leak impact on pressure around the leak location was similar to the pressure response introduced by the PRV, then, the approach could have failed in localising the leak. Therefore using a hydraulic solver which allows such functionality can be more beneficial.

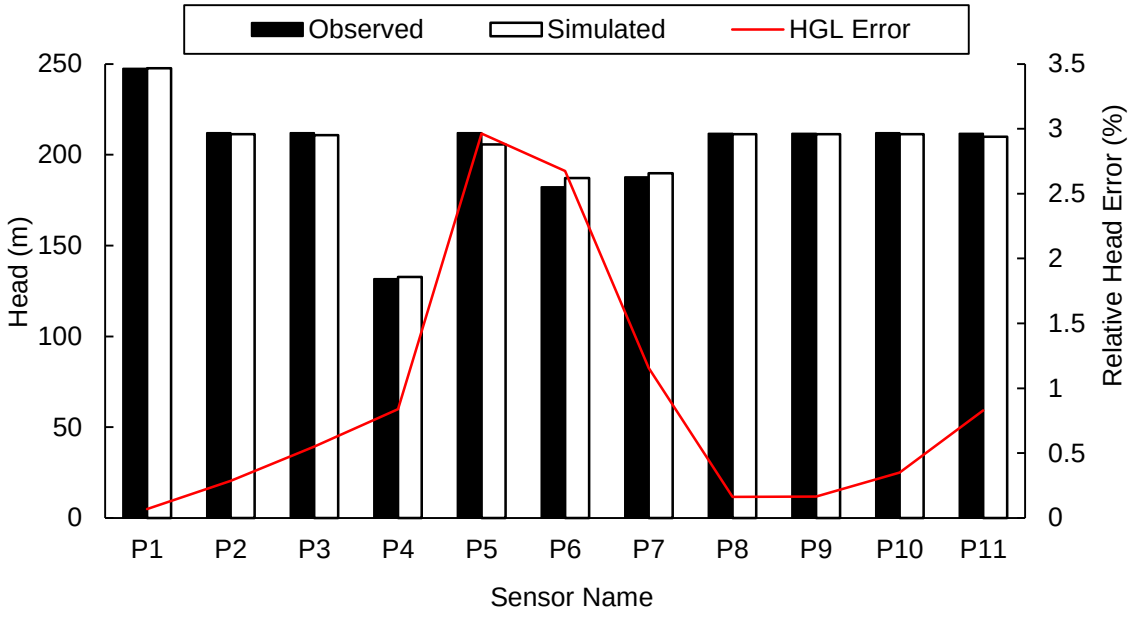

Figure 5.20. Comparison of the observed and simulated heads for the optimised leakage detection solution during MNF at hour 4.

5.3.2 Case Study R2

5.3.2.1 System Overview

Another implementation of the CM and LIM procedures was performed on a larger and more complex UK system (Figure 5.21). The purpose is the same, to eventually detect and localise a real leak event. The DMA is fed by a single source and serves a rural area of approximately 1,800 properties.

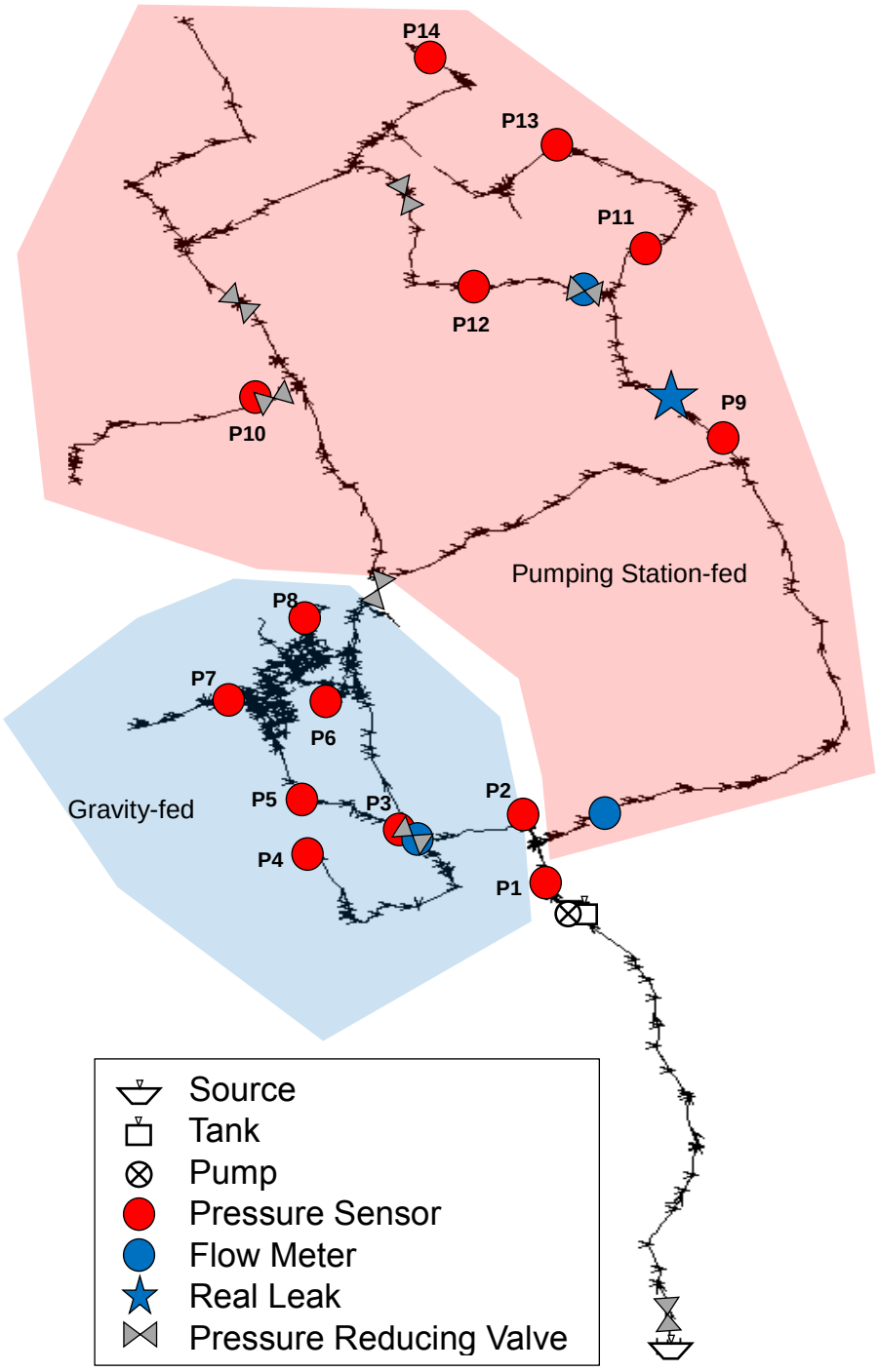

Figure 5.21. The R2 Case District Water System.

Downstream of the inlet there is a tank where water is temporarily stored before it gets distributed across the system. The water is distributed to the customers in two ways. The lower elevation part of the DMA is suppled through gravity, based on the available head from the tank. A pump is operated in the higher elevation part to boost the head and supply required demand. The net flow downstream of the tank varies normally between 3.6 l/s at MNF and 11.2 l/s during the morning peak demand. This represents the combined consumption with both ways of water distribution. The hydraulic model is composed of 940 nodes, 265 valves (not shown in the presented layout) and 702 pipes of total length 48km. From the total number of pipes 41% (287 pipes) are Asbestos Cement (AC), 10% (71 pipes) are Cast Iron (CI) and 49% (343 pipes) are Polyethylene (e.g. PE, MDPE, HPPE). The DMA is monitored for pressure by sensor devices installed at the inlet and at an additional 14 inner locations downstream of the tank. Furthermore, there are flow metering devices at five locations downstream of the tank. This includes the pipe that supplies the gravity-fed part of the DMA and the pump. In total, the DMA contains six PRVs due to the large range in elevation with the lowest point being at 28m and the highest at 186m. The pressure in both parts varies similarly between 26 and 108 metres of head, with an average DMA pressure of 50.3m.

5.3.2.2 Available Data

The raw dataset of observations comprised of pressure and flow readings that were collected for 28 days during the period between November 13 and December 9, 2017 (Figure 5.22). The LIM process used data from a historical leak, which has been detected and localised by the utility. On November 22, 2017 at 17:15 hours, a leak started within the sub-area supplied by the pump and lasted for around twenty days. The leak became visible and was reported by the public in the afternoon of November 30. The leak was fixed on December 12. The exact location where it was found and repaired by the leak detection engineers is indicated in Figure 5.21. A systematic and rising flow difference between 2.25 l/s and 3.05 l/s was observed at the MNF throughout the seventeen days when leak was running, from 1.50 l/s to 4.55 l/s. Moreover, the leak caused a significant and increasing head loss around the leak area of between 6m to 9m of head (Figure 5.22).

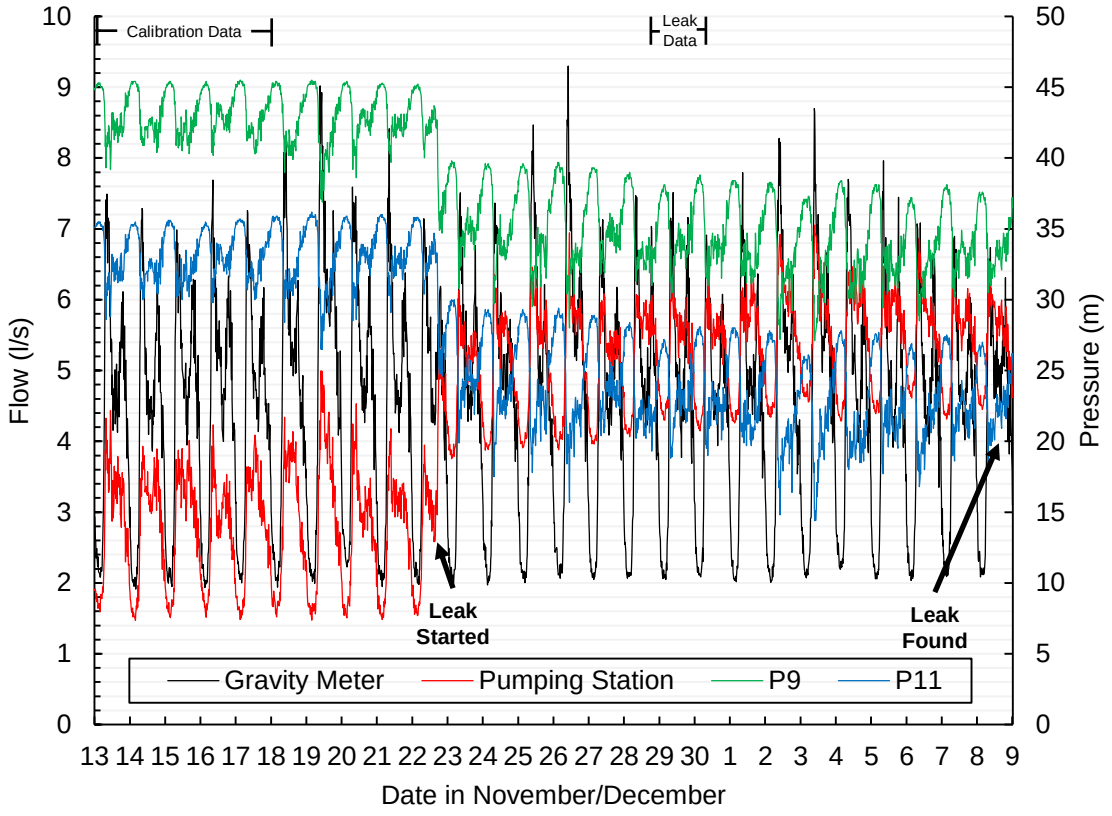

Figure 5.22. The Field Record with the measurements used for model calibration and for finding the leak.

Again, the WDN hydraulic model was calibrated before the leak localisation procedure was applied. This was needed to simulate the average day hydraulic behaviour of the system. A dataset of recorded raw observations for five days (between November 13 and 18, from midnight to midnight) before the leak happened were used to calibrate the model using the CS. The observations for the calibration part involved the time series of flows into the system and the pressures at 14 locations. Similarly to R1, a de-noised calibration dataset of the average day was produced from the raw observations, following the DPP stage. Therefore a total of 96 calibration measurements over a 24-hour period were used for the calibration of the average day hydraulic model. On the other hand, the observed data for the leak detection and localisation part using the LIM involved the 96 flow and pressure observation datasets from midnight of November 29 to midnight of November 30. This was twelve days before the leak became visible to the public.

5.3.2.3 Calibration Method Implementation

5.3.2.3.1 Data Pre-Processing

The consumption during every time step of each day in the raw observations varies within ±4.9%, with abrupt changes in the flow observations. This is because the consumption is mainly domestic and thus, a “noisy” pattern is an expected outcome. During the DPP stage the raw datasets for pressure and flow were analysed to create a smooth average profile for each device which served as the observations within the CS. The datasets for five days from each flow metering device were converted to a 24-hour dataset of the average conditions for subsequent model calibration. The calibration dataset was created within 20 minutes. The outcome of the DPP stage is illustrated in Figure 5.23.

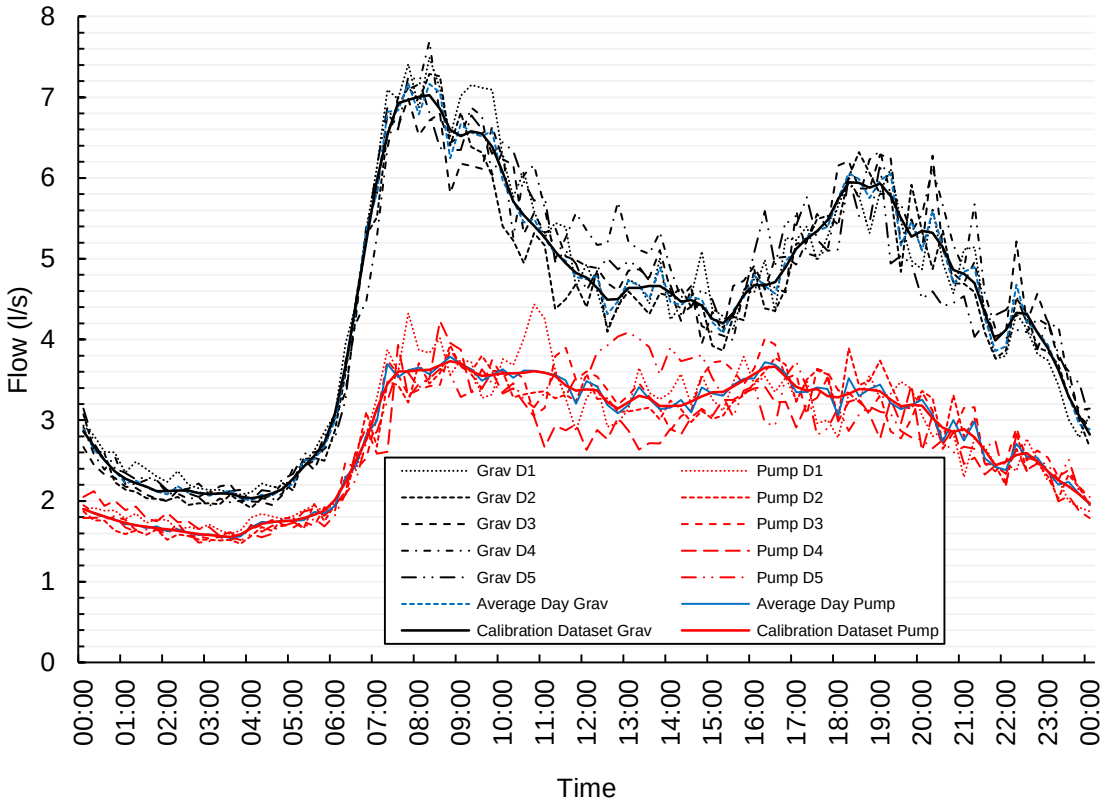

Figure 5.23. Comparison of the resulting average day dataset and the calibration dataset after the DPP stage, relative to the raw observations.

5.3.2.3.2 Profile Calibration

The DPC was carried out by optimising the 96 demand multiplier coefficients for the domestic pattern. The aim was to match the simulated flows at the metered locations with the observed ones. It also considered the net inflow into the DMA

resulting from the combined gravity-based demand and the demand supplied from the pumping station. Again, the pattern for simulating the background leakage profile was not considered in the calibration process, assuming that it remained constant relative to the day before the leak started. This does not impact the results from the LDL methodology, however, it can affect the calibration of the WDN model, as the change in background leakage m. Before DPC is implemented, the simulated net flow showed a discrepancy of at least 0.5 l/s and up to 2.5 l/s. However, following DPC a good match was achieved with the observed net flow with an average error range over the 24 hours of ±2.5% (Figure 5.24). This created an improved representation of the average demand conditions in both the gravity-fed and the pumping station-fed areas, within five hours. Then, at PPC the setting of five of the system PRVs was considered for calibration. Additionally, the pump speed profile was also considered for calibration. The pattern multipliers were calibrated for each time step, which ultimately led to a calibrated hydraulic model at macro-level. The PPC process took another four hours of computational work. The macro-calibrated model was the fine-tuned in the CC stage.

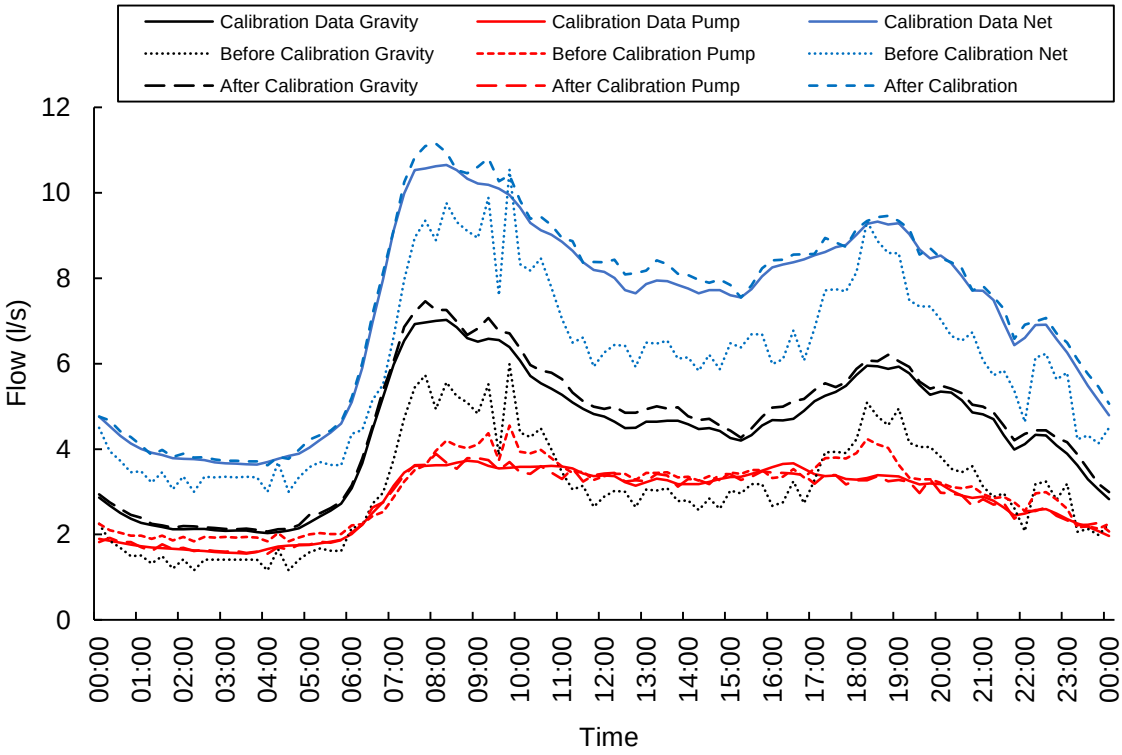

Figure 5.24. Comparison of the updated hydraulic model before and after Demand Profile Calibration as well as relative to the observations.

5.3.2.3.3 Component Calibration

In the CC stage the search space was initially reduced, based on a similar method as in Case Study R1. Overall, about three hours of analyses were required for both valve and pipe components. After the SSR stage only 9% of the valve locations, i.e. 23 out of 245 components, were considered to identify a maximum of 5 closed valves within the gravity-fed area. In addition, 3 out of the total of 26 pipe groups (one AC group and two CI groups) were optimised for their roughness values. This comprised 15% of the WDN pipes, i.e., 109 out of 702 pipes in the WDN model. Following on from the CC stage, optimisation detected four closed valves in that part of the WDN (Figure 5.25). Two of those valves were identified at an incorrect status (Figure 5.25). One of the valves was unknowingly found to be closed and the other one open. This means that the flow path leading towards the sensor device P8 was simulated incorrectly in the hydraulic model. Before calibration the simulated output indicated that water was supplied to that part of the WDN after passing through the PRV, whereas in reality it is supplied without any pressure management. The status of these valves was later checked and verified in the field following the modelling analyses.

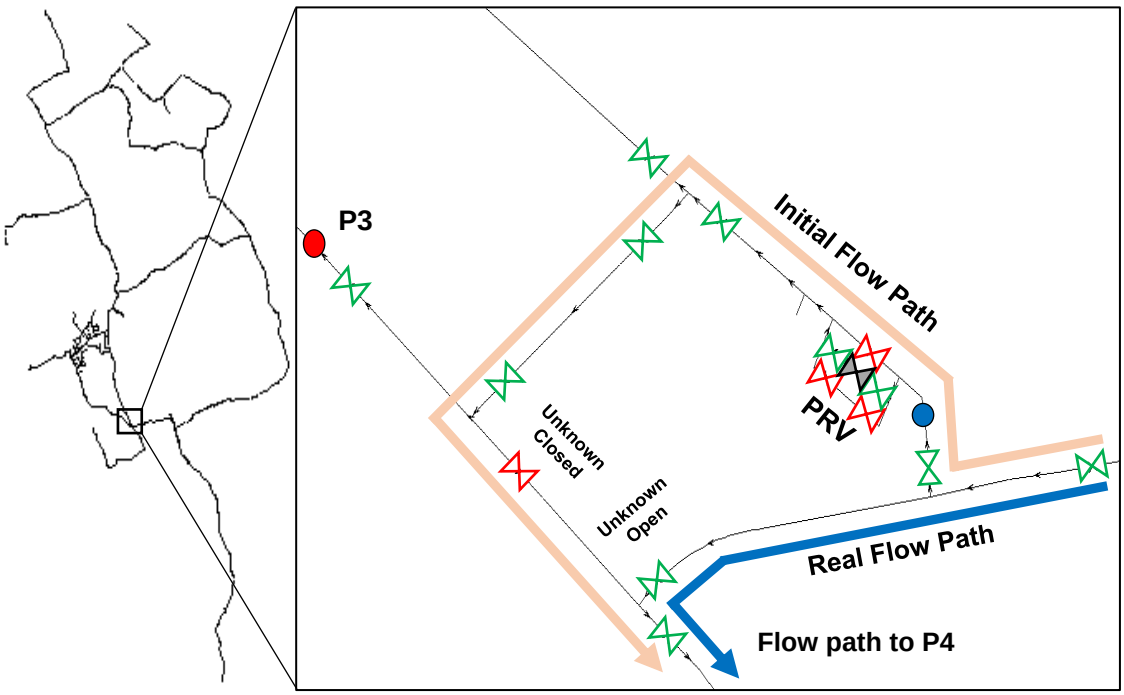

Figure 5.25. The detected correct status of the identified valves following the model calibration.

A comparison of the simulated heads at the sensor locations before and after calibration is given in Figure 5.26. The figure also shows the observations during the hours of MNF and the morning peak, i.e. at 03:30 and 08:00 hours, respectively. The initial model error was reduced to less than 2% across all sensors during both the MNF and the morning peak. The only exception is sensor P8, where a large difference of more than 11% is observed. This was caused by the two valves with incorrect status.

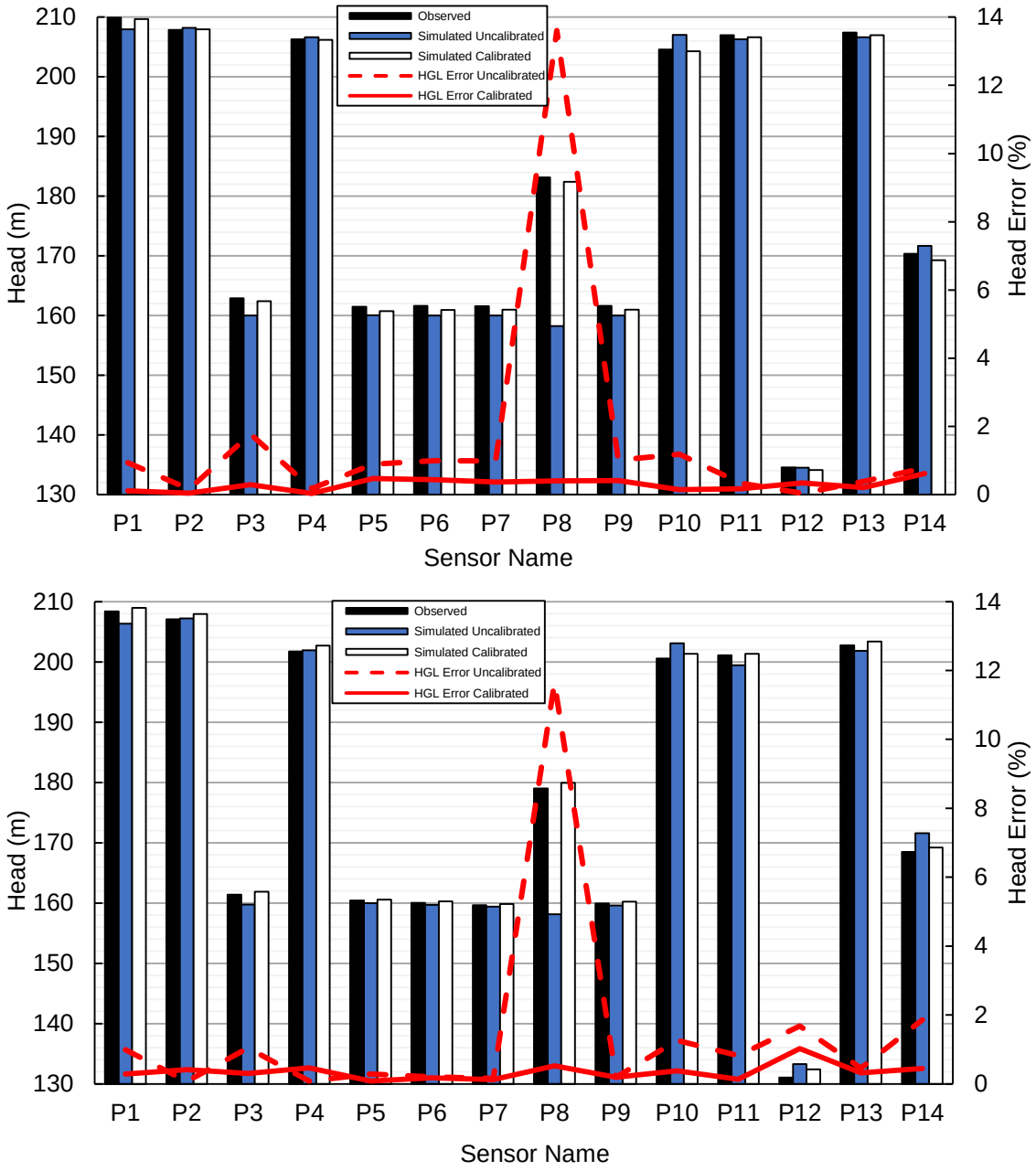

Figure 5.26. Comparison of the observed and simulated heads for the uncalibrated model and optimised calibration solution during MNF at hour 3 (top) and the morning peak at hour 8 (bottom).

Compared to the starting uncalibrated hydraulic model, the calibrated model matched the pressure observations within an error range of ±0.5% across all sensors. This is the calibrated model, which was used for leak localisation purposes. However, the calibration criteria were not fully met. The error ranges for the pressure specific criteria, i.e., ±0.5, 1 and 2 metres, were partially met for 55%, 85% and 96% of pressure data. However, all criteria associated with the simulated head loss were fully met. The reason for missing the pressure specific criteria is associated with the limited modelling capability of the EPANET hydraulic solver. The hydraulic model used in this research for calibration was converted to the EPANET version from the original Synergi software model (DNV-GL, 2018). That version is maintained by the water network operator. EPANET has limited abilities compared to other commercial products and, thus, various settings from Synergi are not supported and can be lost during the conversion process. For example, the profile variation of PRVs, a pump and a tank, which were all part of the presented system could not be fully replicated and combined by the EPANET model during calibration.

5.3.2.4 Leak Inspection Method Implementation

After the SSR stage, which took about two hours to complete, the domain was restricted to 11% of the WDN nodes as the search space for a single leak scenario. A similar weighting scheme as in Case Study R1 was used. This gave an increased weight to the pressure and flow differences during the hours of low demand, i.e., between 00:00 and 06:00. A leak of size 2.65 l/s was detected, which also corresponds to around 25% of the average inflow readings. Based on the detected flow and the MDNL values of the candidate nodes, 20 leak scenarios were simulated, ranging from a single leak of 2.65 l/s up to 20 smaller ones of approximately 0.13 l/s. The fittest scenario suggested a single leak in the WDN. When the leak detection and localisation approach was implemented, the speed setting of the pump was also allowed to vary in order to emulate the response of a real pump. This corresponds to a real situation in which the pump supplies more water and compensates the head loss when a leak happens. The LDL stage was completed within four hours. The optimum solution found a node on the pipe where the real leak occurred while being 150m upstream of the actual location (Figure 5.27).

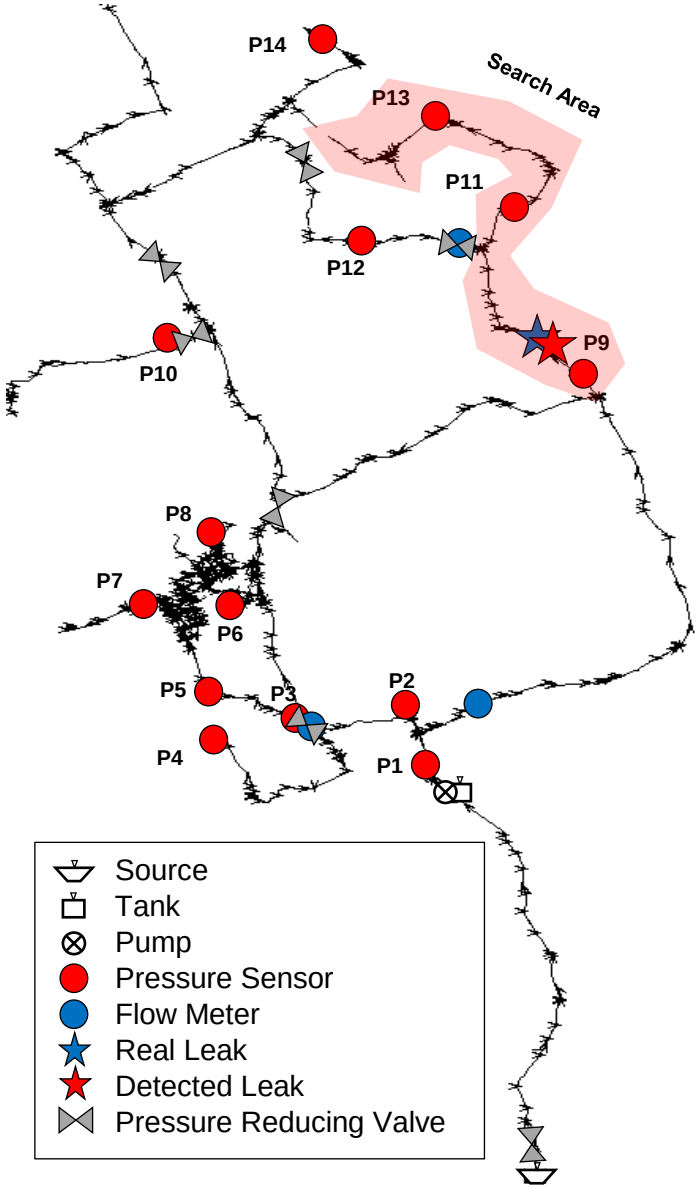

Figure 5.27. Leakage Hotspot Map after the optimisation analysis.

Considering the length of the WDN mains, this represents an error of less than 0.5% (by length). This suggests that the approach was able to localise the leakage area within a street-level distance. It should be noted that the closest model node to the leak was the reported one. Furthermore, the search for a leak was narrowed down to the area between sensors P9 and P11. This corresponds to a maximum search distance of only 10% of the total mains length. The leak was reported to the utility on November 30, thus, the result suggests the approach could identify the leak before it reached the surface and became visible to the public.

The simulated flows using the updated hydraulic model, i.e., including the identified leak, showed a good match with the leak observations over 24 hours (Figure 5.28). Thus, the LDL methodology showed success in both identifying the leak size and location while calibrating the hydraulic model, although the continuous large variability in demand and the rising MNF. The simulated inflow after LDL produced a much better agreement with the observations on November 29 and the following days. Moreover, a very good agreement was also observed with the observations on November 27 and 28 that were not considered in the LDL process. This is evident especially between 00:00 and 06:00 where a higher weight was assigned.

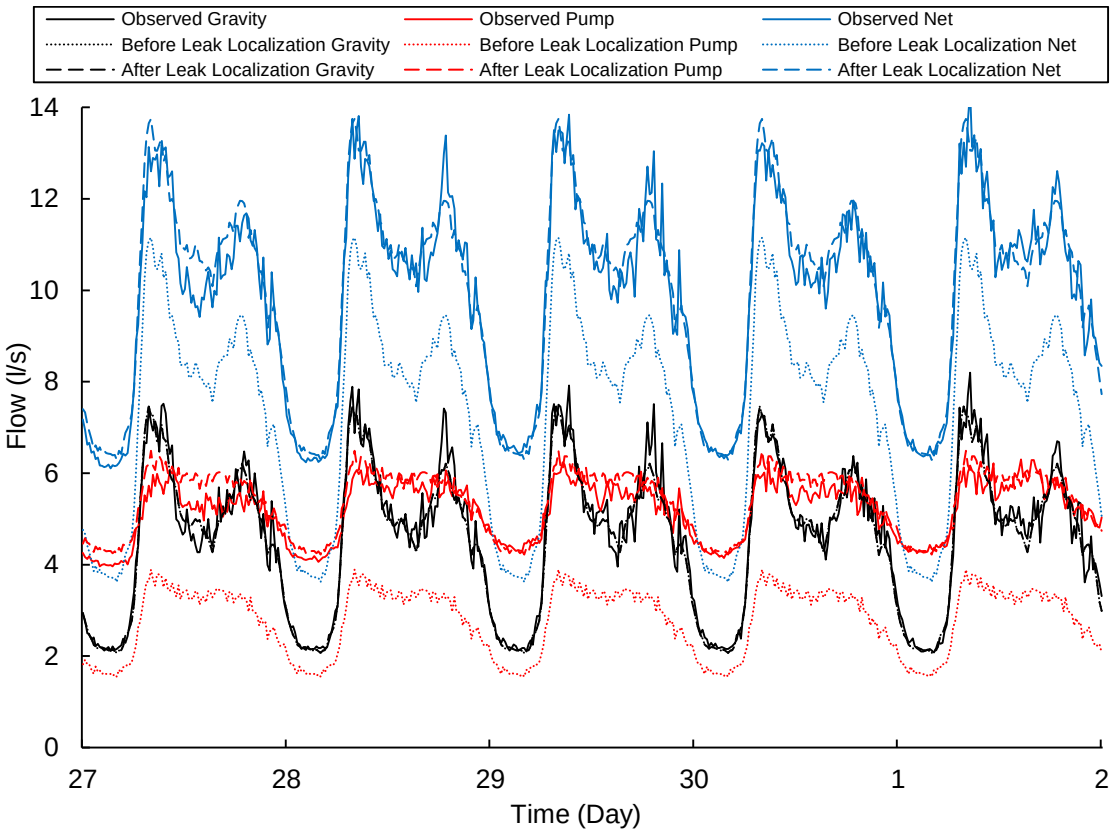

Figure 5.28. The Flow Differences before and after leakage detection.

The comparison of the observed and simulated heads at the sensor locations at 03:30 hours demonstrates the good agreement with observations (Figure 5.29). The error dropped again in the range of ±0.5% across all sensors, which represents a well calibrated model, i.e. within ±1m accuracy relative to the observations.

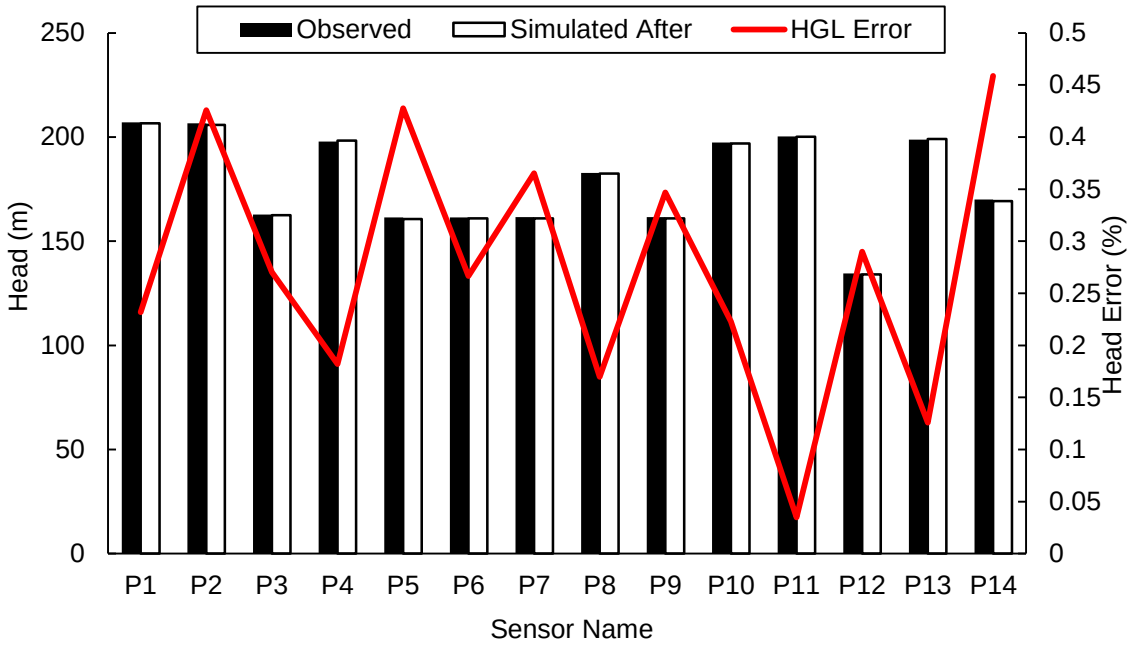

Figure 5.29. Comparison of the observed and simulated heads for the optimised leakage detection solution during MNF at hour 4.

5.4 Evaluation of LIM and CM

5.4.1 Operational Benefits of LIM and CM

The real case applications showed that significant operational savings can be achieved using both the LIM and CM. The use of LIM can reduce the time required to detect and localise a leak, as well as identify the lost volume of water. For example, considering that the leak flow in Case Study R1 was on average 2.25 l/s and ran for a week, the total losses were around 1,400m^3. If the offline approach was implemented using the data of November 8, which would have allowed the leak to be found and repaired by the Leakage Technicians on the afternoon of November 9, there would have been volumetric water savings of more than 70%. Using the same philosophy in Case Study R2, the leak localisation approach could have saved more than 55% of the lost water. This is based on the data used to find the 2.65 l/s leak event, a week before it was discovered and repaired. Except from the volumetric water savings, serious economic impacts can be avoided from costs in fines and insurance payments. Further environmental consequences, such as soil erosion, could also be reduced. For example, in 2011, Thames Water was fined £8.5m for missing their

leakage target (The Guardian, 2017). In addition, a single burst led to insurers to face a £4m payment due to flooding and impacts on transport, homes and businesses (Insurance Times, 2013). Finally, icy weather conditions caused a 4,000% increase in Severn Trent's burst alarms and millions of pounds worth of damage (Severn Trent Water Ltd, 2018). Therefore, the promising potential for a plethora of benefits for the water sector is obvious following the real case validation of the LIM.

Using the CM, the average day hydraulic conditions of the WDN can be accurately represented in the hydraulic model for any DMA and consumption type. The CM considers the demand uncertainty, as a result of fluctuations in consumption, and determines the main flow and pressure profiles in the WDN. Furthermore, any system and data anomalies associated with unknown closed and open valves can be identified, therefore calibrating the flow paths in the WDN model. This not only improves the overall model accuracy, but also the outcome of the leak localisation procedure when implementing the LIM. In both real case studies presented, the starting uncalibrated hydraulic model of the WDN was adjusted to simulate the average conditions of the available measurements prior to the leak occurrence. Interestingly, in both WDNs the valves of incorrect status were detected and corrected. This led to a correct simulation of flow paths in both systems. These were later verified in the field after checking the reported flow routes. With a much improved representation of the WDN behaviour without considering new leaks, any anomaly from a new occurring event could be detected and localised more accurately and allow its earlier pinpointing in the field. The CM implemented on both real-life cases demonstrates the benefits that could be secured by developing an optimisation-based approach rather than pursuing the custom-and-practice methods that have remained broadly the same for 20 years across the UK Water Industry.

The developed search space reduction methodology embedded in both the LIM and CM can highlight the observable and sensitive parts and components of the WDN that can be considered when solving the *inverse problem*. This reduces the number of identified false positive leaks and the incorrectly reported closed or open valves. Consequently, the search effort required to pinpoint the reported

anomalies is minimized. More specifically, during the parameter sensitivity analysis, any pipes and valves where a change in their state and status, respectively, produces a pressure change equal or less than the measurement error are excluded. Therefore the calibration problem is simplified. Furthermore, the MDNL analysis defines a threshold flow value that is detectable based on the sensor configuration. Therefore, smaller leaks that cause local impact (not captured by existing sensors) will require additional measurements. Except from that, the parameter sensitivity analysis provides the water utility an accurate estimate of the level of monitoring a DMA would require. Based on its outcome it can highlight parts of the WDN where more logging or flow metering is necessary for improved WDN observability and leakage control. Combined with externally-based methods used for pinpointing leaks, the LIM provides the opportunity for reducing the time and space required to identify leaks and achieving monetary savings along with higher levels of operational efficiency. In addition, the CM can complement the LIM and contribute to identifying the better starting point for the leak localisation process. Eventually, the synergy between a hydraulic model, an optimisation technique and a systematic search reduction technique, can eventually automate leak localisation and model calibration, resulting in major advances in water network modelling across the UK Water Industry.

5.4.1.1 Computational Benefits of Search Space Reduction

Important benefits in both *inverse problems* of leak localisation and model calibration were secured through the problem simplification. This is achieved by removing unobservable components from the search space. This caused a significant reduction to the number of decision variables. That, in turn, achieves avoidance of unnecessary iterations during the optimisation analyses. The parameter sensitivity analysis restricted the number of components and the range of possible values they can be adjusted during the optimisation analysis. Therefore, the number of solution combinations in leak localisation and calibration procedures was minimized. Eventually, the sensitive and insensitive parameters of the hydraulic model are determined and considered in solving the *inverse problem*.

The MDNL flow analysis restricted the range of possible emitter coefficients to only highly sensitive values for the model fitness. To achieve this, the computations of the forward problem required by the MDNL analysis depends on the number of candidate nodes and the number of emitter coefficients after the user specifies the maximum possible flow. This is represented by Equation 5.1:

$$CM_{MDNL} = NCN\ x\ NEC \tag{5.1}$$

Where: CM_{MDNL} is the number of hydraulic simulations required for the MDNL analysis.

NCN is the number of candidate nodes after Step 1 of SSR.

NEC is the number of emitter coefficients that result after the user specifies the maximum possible flow before the MDNL analysis.

From the calibration point of view the number of detectable valve locations and the detectable pipe groups were established along with their corresponding range of roughness values. The number of computations required to establish the Detectable Valve Locations is equal to the number of candidate valve components represented by Equation 5.2:

$$CM_{DVC} = NCVC \tag{5.2}$$

Where: CM_{DVC} is the number of hydraulic simulations required to determine all the detectable valve locations.

$NCVC$ is the number of candidate valve components after Step 1 of SSR.

On the other hand the Detectable Pipe Roughness Coefficients can be determined based on the number of pipe groups after Step 1 of SSR and the number of the corresponding roughness values given by Equation 5.3:

$$CM_{DPRC} = NPG\ x\ NRC \tag{5.3}$$

Where: CM_{DPRC} is the number of hydraulic simulations required to determine all the detectable pipe roughness coefficients.

NPG is the number of candidate pipe groups after Step 1 of SSR.

NRC is the number of candidate roughness coefficient values that correspond to the candidate pipe groups.

Eventually any change in a parameter value that causes a similar pressure change or less than the measurements accuracy range would practically provide no beneficial use. It will also increase the dimensionality of the search space and, consequently the required computations during the subsequent optimisation analyses. In theory, an over-determined optimisation problem including observable parts of the network as decision variables can be solved with a reasonable accuracy.

Further improvements were achieved by using short optimisation analyses during the SSR stage, i.e., during the SSO step. The power of Genetic Algorithms in significantly improving the objective function value at the initial generations of the optimisation was leveraged for reducing the *inverse problem* dimensionality, while avoiding long and time consuming runs. The number of combinatorial solutions decreased significantly due to the reduced range of possible values for each decision variable in both leak localisation and calibration. During Part I of the LIM in Case Study R1 the maximum emitter for all candidates was reduced by 22% relative to the maximum difference (2.88 l/s) between the observed and simulated system demand. This was achieved only after five optimisation analyses (Population Size: 50, Generations: 30). This is an improvement that none of the previous model-based techniques that use search space reduction achieved. The overall number of hydraulic simulations required to determine the total water losses relies on Equation 5.4:

$$CM_{PartI} = PS\ x\ G\ x\ Z \tag{5.4}$$

Where: CM_{PartI} is the number of hydraulic simulations required to detect the total water losses in Part I of Step 3 in SSR.

PS is the user defined population size for the number of solutions generated by the Genetic Algorithm.

G is the user defined generations number parameter of the Genetic Algorithm.

Z is the user defined number of optimisation analyses (Section 3.5.2.2) for the tested range of emitter coefficient values.

At the end of Part II an approximation of the WDN state was obtained, that involved searching for the location of a maximum one leak. This, reduced the

computations required to explore the search domain of the leak detection and localisation problem according to Equation 5.5:

$$SS_{PartII} = NIndex^{NLeak}\ x\ (K+1)^{NLeak} \quad (5.5)$$

Where: SS_{PartII} is the number of hydraulic simulations required to evaluate any possible solution to the optimisation problem in Part II of SSR and the ultimate leak detection and localisation problem, i.e., $SS_{PartII} = SS_{LDL}$.

$NIndex$ is the number of candidate nodes for any possible leak size.

$NLeak$ is the maximum number leak locations to be identified.

K is the number of possible emitter coefficient values for any potential leak location.

This is based on an objective approach relative to the subjective approach proposed in Wu et al. (2010) and without perturbing the accuracy. If no reduction in the dimensionality of the *inverse problem* in Case Study R1 is undertaken, the search space in the Base Case requires $1x10^{601}$ computations to be explored. This is based on a discretised range of only 10 possible flow values following Equation 5.6:

$$SS_{BaseCaseLeaks} = K^{NIndex} \quad (5.6)$$

Where: $SS_{BaseCaseLeak}$ is the number of hydraulic simulations required to evaluate any possible solution to the base case optimisation problem for leak detection and localisation without any reduction in space.

$NIndex$ is the number of candidate leak location nodes.

K is the number of possible emitter coefficient values for each potential leak location.

These are a lot of computations for the algorithm to converge to an optimal solution, which were avoided using the proposed method. The restriction during the LDL stage to only a number of possible leaks with an additional reduction of the range of possible flow values can contribute to finding leaks earlier. Here, the maximum number of computations for exploring the search space reduced to only 362. Consequently, reducing the search space before solving the *inverse problem* removed 99.99% of false positive solutions, which improves the leak

detection and localisation reliability. To verify this, the leak localisation in Case Study R1 (with the known leak) was solved without a reduced search space (i.e. 100% of WDN nodes). The optimisation ran for 500 generations and a population size of 200, similarly to the case with a reduced search domain. However, the non-reduced case converged to an incorrect solution. The number of maximum possible closed valves in the WDN before model calibration were determined after a series of scenarios, based on Equation 5.7:

$$SS_{ClosedValves} = VIndex^{NValve} \quad (5.7)$$

Where: $SS_{ClosedValves}$ is the number of hydraulic simulations required to evaluate any possible solution to the optimisation problem in finding the maximum number of closed valves.

$NValve$ is the maximum number of possible closed valves to be identified.

$VIndex$ is the number of candidate valve components for any possible closed valve.

More specifically, the number of computations necessary for the search space exploration in Case Study R2 was reduced to 23^5, in contrast to the 2^{245}, for the Base Case, i.e., without any implementing the SSR stage. This was a reduction of 99.99% in the number of possible solutions and was achieved following Equation 5.8:

$$SS_{BaseCaseValves} = 2^{VIndex} \quad (5.8)$$

Where: $SS_{BaseCaseValves}$ is the number of hydraulic simulations required to evaluate any possible solution to the base case optimisation problem for detecting the status of valves without any reduction in space.

$VIndex$ is the number of candidate valve components.

According to Equation 5.9 the problem dimensionality associated with the pipe roughness calibration groups was minimized to:

$$SS_{PRCG} = RIndex^{NGroups} \quad (5.9)$$

Where: SS_{PRCG} is the number of hydraulic simulations required to evaluate any possible solution to the optimisation problem in finding the maximum number of pipe roughness calibration groups.

$NGroups$ is the maximum number of possible pipe groups to be calibrated.

$RIndex$ is the number of possible roughness values for each pipe group.

This resulted in roughly 12^3 necessary computations as opposed to the 12^{702} required for the base case (i.e., 99.99% reduction), based on Equation 5.10:

$$SS_{BaseCasePipes} = RIndex^{WDNpipes} \quad (5.10)$$

Where: $SS_{BaseCasePipes}$ is the number of hydraulic simulations required to evaluate any possible solution to the base case optimisation problem for calibrating the roughness coefficient of pipe components without any reduction in space.

$WDNpipes$ is the total number of pipes in the WDN model.

$RIndex$ is the number of candidate valve components.

All of the above simulations were carried out on an Intel(R) Core(TM) i7-5600U processor of 2.6GHz speed and 16GB memory. Avoiding such a number of unnecessary solution combinations can vastly speed up the *inverse problem* solving procedure. The reason lies in the less decision variables and the complexity of the problem. For example, in practice, cost effective leaks are often associated with a small number of hotspots. Similarly, although the status of all valves in the WDN cannot be continuously monitored, those with an uncertain status are associated with just a few components. Consequently, the initial population of solutions to the *inverse problem* can begin at a better starting point.

5.4.1.2 Practicality of the LIM

The LIM is also applicable to larger WDNs and can detect and localise multiple leaks. However, its accuracy depends on the type, number and location of available pressure and flow measurements as this can affect the outcome of both the SSR and LDL stages. The method works best when the number of observations is larger than the number of potential leak locations. The SSR stage defines the WDN resolution, i.e., all model components that can be determined during inverse analysis. However, a larger number of sensors will lead to an

increased resolution and therefore a larger search space. On the other hand, more available measurements will result in a more unique pressure and flow signatures and, therefore, improved accuracy. Moreover, with a variety of sensor types (i.e., flow, pressure or noise logging) and data collection based on the optimal sampling design, the process can result in more unique solutions and improvement in the condition of the *inverse problem*.

It is expected that any leak that causes a pressure drop larger than the uncertainty caused by inaccuracy of pressure sensors can be detected and approximately localised. This was shown by the semi-real and real case studies, allowing a reduced search effort and earlier pinpointing by the Leakage Technicians. Based on the outputs from the semi-real case studies, leaks of at least 5% relative to the average system demand can be localised at a street-level resolution. Considering the accuracy of contemporary pressure sensors, in Case Study SR1 the leak J-33 was localised within approximately 170m. Complementary, in Case Study SR2 leaks J-305 and J-183 were localised within 330m and 280m, respectively, as an average distance for all demand fluctuation scenarios using CAL D (Table 5.3).

A point to raise is associated with the automation developed in the LIM. This can contribute to significant search effort benefits, associated with the traditional labour-intensive leak detection and localisation procedures carried out by practitioners. However, it requires a calibrated hydraulic model (see section 2.4.5) and a set of well-behaved data (see section 4.3.2). Another important point is associated with the impact caused by the idealized value of the pressure exponent a (2.16 in Section 2.5.4.2.1) considered by the LDL methodology. To minimize that Zonal Tests can be undertaken (Lambert et al., 2017) for a more accurate estimate of a.

Compared to other inverse methods, an advantage of this approach is that it takes into account the uncertainty in observations, which can minimize the number of non-unique solutions and false positives. Furthermore, it reduces the subjective impact coming from the user's choices and establishes an objective system for finding leaks. This provides a much better starting point for the

generated population of solutions in the subsequent optimisation analyses during the leak detection and localisation procedure.

Nevertheless, a drawback of the presented method lies in the outcome of Part II during the SSR stage, which may lead to an incorrect estimate of the WDN state. This can occur due to the premature convergence of the algorithm caused by the non-exhaustive optimisation analyses. The results suggest that multiple optimisation runs for different leakage scenarios prove beneficial for assessing the number of leaks in the network, although in larger systems that can be time consuming. Another key limitation of the LIM is that it can only work for new leaks that occur after the model has been calibrated. This is because background leaks already occurring in the DMA are masked in the calibrated model. Furthermore, a large number of sensors may be necessary. However, the cost of hydraulic sensor technology has reached the level that enables their large scale deployment in a WDN. Finally, due to the large number of simulations necessary to reduce the search space and converge to an optimum solution a real-time implementation of the LIM is currently not likely. From the presented case studies it is expected that a leak can only be identified at least 3 hours after it has happened (1 hour for SSR stage and 2 hours for LDL stage), thus, allowing only near-real time applications. However, this depends on the size of the hydraulic model and the ability of the optimisation technique to evaluate solutions. In addition, on the chosen number of optimisation analyses, the population size and the number of generations specified. On the other hand, with improving computing power or the scalability provided by cloud computing this can be significantly reduced.

5.4.1.3 Practicality of the CM

Similarly important is the type, number and location of available measurements for SSR and CC stage of the CM. However, for an improved model calibration it is strongly suggested that an increased number of flow measurements is used and at least one sensor every approximately 200 properties. This is because a change in the status of a valve, or state of a pipe, which is not located on a key flow route that conveys a significant fraction of the global demand, would cause very small changes in pressure. The starting model from Case Study R2, which

was calibrated based on flow readings from five devices led to a much more accurate leak localisation, as opposed to the starting model of Case Study R1, which only used the inlet flow. The detected leak in Case Study R2 was reported within a ±0.5% error (by length) relative to the 2% error in Case Study R1, although the network was larger, more complex and the pressure logging density was the same. In addition, the offline CM demonstrated the ability to match the simulated outputs in pressure and flow with the observed ones, while identifying valves with uncertain status. An important step in modelling the correct WDN hydraulic conditions was associated with the use of a well-behaved calibration dataset, following the DDP stage. Through the exploitation of all available measurements and the use of a de-noising technique the large demand fluctuation in both real cases was mitigated. This allowed to determine an accurate representation of the average hydraulic conditions. The Case Study SR2 which examined various scenarios of demand uncertainty and calibration, suggests that the CM can accurately calibrate any DMA type characterised by any consumption type. The CM is also an automated and objective procedure, which only requires a set of well-behaved raw measurements (see section 4.3.2) and a hydraulic solver combined with an optimisation technique. Therefore the time consuming trial-and-error adjustments in model parameters based on subjective decisions of the modelling engineers can be eliminated.

A significant drawback of the CM lies within the capabilities of the EPANET hydraulic solver used in this research after its conversion from the original Synergi version. For example, in Case Study R2, it was impossible to calibrate the part of the WDN upstream of the tank as EPANET cannot support any functionalities related to the profile behaviour of the tank water level. Thus, 57 pipes and 23 valves, i.e., 8% and 9% of WDN components, respectively, were not considered in the calibration procedure. In addition, there is no functionality to represent the profile behaviour of a changing PRV setting during the 24 hour simulation. However, a separate code was developed so that a time-step wise calibration of the PRV settings was achieved, similarly to the pattern multiplier calibration, used during the Profile Calibration. Although calibrating each step separately is more time-consuming, the missing functionality in EPANET is approximated and the resulting hydraulic model becomes more accurate. Finally, the WDN hydraulic

models used in the presented case studies were converted from an original Synergi version, maintained by Severn Trent Water, to an EPANET version. Due to this conversion and the limited hydraulic modelling ability of EPANET, many of such settings were lost. Even though, this issue does not invalidate the calibration methodology as a more accurate hydraulic model was still achieved, all of the above reasons suggest that the CM is applied to simpler WDNs for an improved accuracy. The practical use of the CM can be largely improved using a hydraulic solver where these functionalities are accessible, i.e., using a commercially available solver, or in an upgraded version of EPANET (e.g., EPANET 3).

5.4.1.4 Model Calibration and Data Challenges

A key requirement for accurate leak localisation using the LIM and the message from this work is the need to use a well calibrated model of a WDN. However, the traditional methods of calibration, i.e. adjusting pipe roughness coefficients and nodal demands through trial-and-error, have a low chance of meeting the even stricter calibration criterion now required for operational WDN modelling. As demonstrated by both the semi-real and real case studies, a calibrated model contributed to a more accurate leak localisation and improved model quality. With the advent of cheaper telemetry and monitoring devices, there are opportunities to further exploit the information captured for estimating decision variables with satisfactory confidence. More data is required due to the increased complexity of large urban WDNs and the ill-posedness problem associated with WDN modelling. Although the semi-real case SR1 studied a relatively small WDN and used data from the traditional sensor deployment density of 1 per 200 properties, the uniqueness problem was still an issue. This is because the mean pressure change was similar at a smaller leak size. Combined with the effect of noise, various combinations of solutions led to equally fit solutions and eventually to an incorrect localisation. Even using a perfect dataset, at a 5% leak the pressure changes were too small to produce a distinct “signature” that could be used to correctly localise the leak. Significant improvements in the accuracy of the model calibration and leak localisation process can be secured with the inclusion of additional flow measurements captured from key flow routes in the WDN. Furthermore, the impact caused by small unknown leaks, or the local effect caused by unknown closed/open valves, can be often insufficient to allow

detection. An unknown closed/open valve that due to data anomalies may have been left open/closed in the calibrated model, can result in false positive leak detection. This may be associated with incorrect pipe group roughness values that also exist in same flawed models. Therefore an increased number of flow measurements can facilitate the detection of valves with unknown status.

Current WDN models are calibrated to a standard that assumes that simulated model fits observed pressures within ±1m. Thus, hard-to-find leaks and topological anomalies, i.e., faults that cause a pressure change of less than a metre, can remain undetected due to small head losses. It is time to move towards a stricter calibration criterion of ±0.1m, equal to the accuracy of field pressure transducers. This will then be able to provide far superior calibration with increased opportunities for more successful detection of previously undetected model anomalies. This will, therefore, allow mathematical optimisers to more clearly distinguish between leak induced head losses and those arising from other faults resulting from theft, unknown status valves, or incorrect pipe sizes and roughness values. The results in Case Study SR2 suggest that the calibration dataset quality can affect the model fitness in accurately localising leaks. Although all tested scenarios were calibrated within ±0.5m, i.e., better than the desired model performance criteria, demand fluctuation severely affected the distance and size of the reported leaks as opposed to the actual locations and flows. This relies on the DMA type, i.e., urban or rural, and the consumption type, i.e., domestic, industrial, etc. A calibrated hydraulic model of a DMA with large fluctuations in demand may include additional noise, which can mask small leak events and hide the impact caused by unknown closed or open valves. Eventually, the level of effort required to calibrate a WDN hydraulic model and the desired accuracy level relies on its intended use. For the *inverse problems* of leak detection and localisation and model calibration, which are interconnected in establishing the model quality, the set of model performance guidelines should consider demand uncertainty.

Smoothing the raw measurements of a highly variable demand during the DPP stage of the CM proved beneficial in representing the main profile of flow and pressure in the real WDN. This contributed to more robust solutions during the

leak localisation and elimination of false positives. Therefore, choosing the right calibration dataset, based on the DMA and consumption type is key for accurate modelling of the WDNs. On the other hand, with the aim of matching simulated outputs with the observed ones, the subjective calibration dataset currently used across the UK Water Industry has a low chance of providing accurate leak localisation in systems with large demand uncertainty. Thus, calibration of WDN hydraulic models should consider the DMA and consumption type when choosing the ultimate right calibration dataset for modelling the average day hydraulic conditions of the system.

5.5 Summary and Conclusions

In this chapter, several case studies were carried out with the following main objectives: (1) to verify the leak inspection method and the calibration method proposed in Chapter 4; (2) to demonstrate and analyse the effects of the number, type, accuracy and locations of the available data in the accuracy of leak localisation and model calibration; (3) to compare various datasets used for calibrating hydraulic models which are used to find leaks in WDNs. Two offline model-based methods for leak detection and localisation and model calibration applications were presented. In both methods the size of the *inverse problem* is reduced before solving it. The LIM uses head and flow measurements to detect and localise leaks in WDNs, while the CM to accurately calibrate the WDN and identify any unknown closed or open valves.

The LIM has been tested on four UK networks using artificial and real leak data, respectively. The desktop cases demonstrated the advantages and limitations of this method, depending on the quality of data and model calibration. The method was then validated by detecting a real leak event from two UK DMAs, proving that the area for finding the leak on the ground can be reduced to about 10% of the WDN (by length), which could have led to savings in water of more than 50%. In both semi real and real cases presented, the optimum solution was preserved in the final search space. The two systems were evaluated in terms of operational, computational and practical benefits. A discussion was also provided on the major issues with the traditional calibration of WDN models and the ill-

posedness of the calibration problem. In the final chapter of this thesis an overall summary is provided and the main conclusions are drawn based on the key findings, along with suggestions for further avenues resulting from this research. The main conclusions from this chapter are as follows:

- The search space reduction stage can be used to formulate a simplified version of the initial *inverse problem,* comprised of observable and sensitive parameters, while defining the number of its possible solutions. This achieves avoidance of unnecessary iterations during the optimisation analyses and reduces the number of identified false positive leaks and the incorrectly reported closed or open valves.
- The LIM can be used for the automatic detection and localisation of leaks in rural and urban WDNs at a street-level resolution. Consequently, the search effort required to pinpoint the reported leaks is minimized. It is expected that any leak that causes a pressure drop larger than the uncertainty caused by inaccuracy of pressure sensors can be detected and approximately localised by the LIM.
- The CM can be used as an automatic and objective procedure for the calibration of average day models and the identification of valves with uncertain status, where only highly sensitive parameters to the model fitness are adjusted. Therefore the time consuming trial-and-error adjustments in model parameters based on subjective decisions of the modelling engineers can be eliminated.
- The calibration dataset and the demand uncertainty has a direct impact on the hydraulic model quality and the accuracy of leak localisation, even the calibration criteria are met. The availability of a larger number of flow measurements results in more precise detection of incorrectly modelled valves and improved model calibration. This can ultimately contribute to a more accurate leak localisation process.

CHAPTER 6

Summary, Conclusions and Further Work Recommendations

6.1 Introduction

WDNs are critical infrastructures that increasingly large numbers of people rely on daily. As population grows, the demand on these critical infrastructures also increases. The need for accurate WDN modelling then becomes vital to ensuring reliable and adequate rehabilitation and operational solutions. Sensing technology has advanced to the point that the deployment of dense networks of low-cost pressure and flow sensors is now feasible. When placed at optimal locations, the increased density and availability of measurements from these devices can improve the observability of the system, i.e., the length of the network where the model parameters can be determined. This can enable more accurate leak detection and localisation, improved model calibration, and consequently, more proactive management of WDNs.

The primary aim of this research was to develop novel approaches that simplify the deployment of optimisation techniques for solving *inverse problems* in WDN modelling applications. The particular applications considered are associated with leak detection and localisation, as well as model calibration. This is so that eventually the leak localisation process by means of hydraulic modelling becomes more reliable and faster, and the model quality improves. Since both *inverse problems* are non-linear, non-convex and multi-modal, a robust optimisation technique is required. A genetic algorithm has been commonly used with a WDN hydraulics solver engine to accomplish this task. Pressure and flow data collected at key points in the system are used to estimate the model parameter values. A summary of the results of this research is presented in the following sections. The contributions and conclusions drawn from this thesis are

presented next. Finally, some recommendations for further work are made, based on the results presented in this thesis.

6.2 Summary of Present Work

A novel search space reduction technique has been presented in this thesis as part of a proposed decision-support framework for solving inverse problems in WDN modelling. The technique has been developed so that the dimensionality of an *inverse problem* in WDN modelling is minimized. Its main characteristics are as follows:

(1) Three types of decision variables are considered: (a) the node emitter coefficient values in the WDN hydraulic model used for leak detection and localisation; (b) the status of valve components used in model calibration for detecting data anomalies in the hydraulic model; (c) the roughness coefficient value of pipe components used in model calibration for estimating flow velocities in the real system.

(2) For each decision variable type the search space reduction is performed in three steps: (Step 1) Inverse Problem Simplification; (Step 2) Parameter Sensitivity Analysis; and (Step 3) Search Space Optimisation.

(3) Prior information, including expert knowledge of the system, and a number of assumptions are employed to simplify the *inverse problem* in Step 1.

(4) The pressure sensor number, locations and accuracy range are considered and through a sensitivity analysis of parameter values the observable parameters and parameter values are identified in Step 2.

(5) A number of optimisation analyses are performed in Step 3, based on various scenarios for the system state considering the available observations to, ultimately, restrict: (a) the number of candidate leak locations and the range of possible flow values; (b) the number of candidate valve locations with uncertain status and; (c) the number of candidate pipes and the range of possible roughness values.

To improve the reliability and timeliness in localising leaks and the overall model accuracy two methods were developed that integrate the search space reduction technique, so that the *inverse problem* dimensionality is reduced.

6.2.1 Leak Detection and Localisation

A novel Leakage Inspection Method for detecting and localising leaks was proposed with the following main characteristics:

(1) A calibrated model is a necessary prerequisite for more reliable leak localisation results.

(2) The leak localisation is formulated as a constrained optimisation problem of weighted least square type with a reduced decision search.

(3) The Leakage Inspection Method detects and localises leaks in DMAs, based on systematic search space reduction, where unknown leaks are simulated as node emitter coefficients. This has been tested and validated on semi-real and real case studies.

(4) The method integrates two stages: (a) a Search Space Reduction stage, where the number of candidate leak locations and the range of possible flow values is reduced; and (b) a Leak Detection and Localisation stage, which considers the reduced set of decision variables within an optimisation analysis to find the size and location of leaks.

6.2.2 Hydraulic Network Model Calibration

A new, consistent and unbiased Calibration Method has been developed that determines the state of internal pipe roughness values, the setting/speed of any PRV/pump and detects the correct status of any incorrectly modelled valves. Its main characteristics are as follows:

(1) The calibration is formulated as a constrained optimisation problem of weighted least square type with a reduced number of decision variables. A systematic search space reduction is applied to reduce the problem dimensionality.

(2) The Calibration Method involves four stages: (a) Data Pre-Processing stage to generate the calibration dataset; (b) a Profile Calibration stage for macro-level calibration of the WDN model; (c) a Search Space Reduction

stage for reducing the calibration problem size and; (d) a Component Calibration stage for micro-level calibration. The system has been tested on semi-real and real case studies.

(3) The Data Pre-Processing stage generates a well-behaved problem dataset from the raw sensor data. Those datasets are comprised of measurements that represent the average day hydraulic conditions in the WDN. This is performed by implementing two steps: (Step 1) Average Day Profile Dataset Generation and (Step 2) De-Noised Dataset Generation.

(4) During the Profile Calibration stage, the profile of any demand category and setting of any model component that affects the WDN model flow and/or pressure, is calibrated using inverse analysis.

(5) At the Search Space Reduction stage, the pipe and valve components are analysed and the search domain is reduced before the calibration of the WDN model.

(6) At the Component Calibration stage, an optimisation problem is formulated and solved where the pipe group roughness coefficients and status of valves are fine-tuned to calibrate the WDN model and detect any incorrectly modelled valves (i.e., closed or open) in the WDN.

6.3 Research Contributions

6.3.1 Inverse Problem

The present work has made a number of contributions to the field of Inverse Problem solving, summarized as follows:

- **Development of a novel search space reduction methodology for solving leakage related inverse WDN problems:** One of the main contributions of this thesis is the method that incorporates prior information for a number of components to alleviate difficulties with and improve the condition of the *inverse problem* associated with leakage detection and localisation, and model calibration. This minimizes the computational effort and improves efficiency of the corresponding optimisation problem. Ultimately, this increases the ability to perform robust search and get as closely as possible to the global optimum.

- **Synergistic use of expert knowledge, the available data type, number and locations, as well as a search method to define a reduced number of possible solutions to the leakage related inverse WDN problems:** Another main contribution of the work done in solving an *inverse problem* is the definition of the search space. The use of prior information and expert knowledge of the system has contributed into more practical and realistic solutions. Based on the type, number and location of the available data a threshold minimum and maximum value is defined for each candidate parameter. This is done for both the leak detection and localisation problem and the model calibration problem. The accuracy range of the pressure sensor devices determines the detectable parameters and defines the decision variable threshold (bounds) values. This provides the search process a better starting point and in minimizing the risk of premature convergence and increases the confidence in the possible solutions.

6.3.2 Leak Detection and Localisation

A number of contributions were made to the field of leakage management:

- **The development of new offline simulation-optimisation method for leak detection and localisation based on search space reduction:** The proposed Leak Inspection Method uses a search space reduction method before applying leak localisation so that the computational effort in finding the leak(s) size and location(s) is minimized. Ultimately, the leak search process starts at more targeted locations, where a reduced area and effort is required to pinpoint leak/burst events.
- **A novel use of sensitivity analysis methods for determining the minimum detectable leakage flow and highlighting unobservable parts of the WDN:** The parameter sensitivity analysis step of the search reduction method can be used to define the threshold detectable flow, for every candidate leak location in a WDN. As a direct outcome the analysis determines the parts of the WDN where more or different data are necessary to determine the WDN state.

6.3.3 Model Calibration

Finally, the contributions to the field of hydraulic network model calibration are as follows:

- **Development of a new calibration method for average day models based on data pre-processing and search space reduction:** The proposed calibration method combines a data pre-processing for generating an accurate dataset of the average system conditions, and a search space reduction method before the model calibration is attempted. This allows for a more efficient and effective calibration of WDN components
- **Exploitation of all available sensor data for calibration, and development of an effective data pre-processing method for determining average day patterns:** The proposed data pre-processing method enables detecting and correcting erroneous values in sensor data, while capturing the general pattern of the average day hydraulic conditions, using all available field measurements. Therefore, the whole raw dataset can be effectively converted to fit-for-use data for model calibration.

6.3.4 Water Distribution Network Modelling

The present work has also made a number of contributions to the field of Water Distribution Network modelling:

- **Generalization of the inverse analysis methodology for straightforward application to any District Metered Area:** The development of both methods for leak detection and localisation and model calibration has considered that the final application must be able to be applied to any District Metered Area. This is supported by the successful application of both methods on real networks of different size and complexity, as well as various scenarios for the leak size(s) and locations.
- **Development of a modular software suite that includes the developed methodologies:** The proposed methodologies developed are being encapsulated in a MATLAB GUI, providing an automated system for

identifying the reduced parameter set and for allocating the optimised parameters. Therefore minimal effort is required by the Modelling Engineers.

6.4 Research Conclusions

A hydraulic network model can only be used with confidence if it reveals the real behaviour of the water distribution system with reasonable accuracy. Such models can provide adequate and automatic support in providing solutions associated with operational work at distribution mains level, such as finding leaks. Solving the *inverse problem* in WDN modelling is usually a very complex task that requires field measurements. Those are used to determine the unknown causal parameters, relevant to model calibration and leak localisation purposes. Rather than using some trial-and-error approach, an optimisation problem can be solved. However, to develop a well-posed formulation of the problem, the solution must be both unique and stable to noise in data. In population-based algorithms, such as a Genetic Algorithm, the quality of the initial population plays an important role for their performance. For example, if the initial population contains some good solutions, the algorithms converge quickly. This research showed that several improvements can be made to condition an ill-posed problem and to find better starting point for optimisation analyses. The improvements are achieved by reducing the search space while reducing the risk of missing the optimum solutions.

The key messages from this research are as follows:

- The Leak Inspection Method is able to detect and localise leaks effectively in a WDN using a calibrated model of the system. The approach is flexible and can be applied to networks of varying sizes and characteristics. This was demonstrated on two real case studies where the leaks were successfully identified. The method ensures that a good starting point is found, which achieves high computational efficiency leading to early and reliable leak localisation. When compared to the Wu et al. (2010) approach, which introduces a subjective element to the search for a solution, the LIM considerably reduces the risk of eliminating the best

solution and minimizes the subjective element of the leak localisation approach. In addition, the novel procedure suggested here minimizes the maximum number of computations needed before implementing the leak localisation. This improves the uncertain number of iterations required by other methods, such as Nasirian et al. (2013).

- The Calibration Method can identify throttle valves with uncertain status, determine the state of internal pipe roughness and the setting/speed of any PRV/pump. Furthermore, it can simulate the average day hydraulic conditions and ensure that a good starting model is provided for leak detection and localisation. The calibration problem is solved automatically by an optimisation-based method, therefore, improving upon the custom-and-practice WDN modelling procedures followed by practitioners. However, it is not reasonable to expect that the calibration problem will be solved using a completely automated procedure.
- The accuracy of leak localisation and the effectiveness of model calibration is affected by the performance of the LIM and CM. For an accurate leak localisation a well calibrated model is necessary, as its ability to identify leak events improves with more accurate predictions. Similarly the hydraulic model predictive quality improves when a leak is found and the model is updated after the leak is fixed. Furthermore, collection of additional pressure and (especially) flow data, can bring multiple benefits to localising leaks and improving hydraulic model calibration. This is due to a unique “signature” being created on the recorded data for each situation in the real system, which helps to identify any anomalies more accurately.
- The offline model-based methodology can improve the *inverse problem* condition. The improvement is achieved after a systematic reduction in the number of decision variables and the range of possible values by considering the error in the available data. However, the outcome of the search space reduction method depends on the type, number, accuracy and the topological placement of sensor devices. Therefore, the *inverse problem* condition and dimensionality is interconnected to the sensor placement design.

- At the beginning of any *inverse problem* in WDN modelling, it is important to develop a detailed understanding of the system hydraulics and water consumption. Furthermore, it is advantageous to identify clear and realizable expectations for the solution to the problem. Any information available that is directly or indirectly related to the aim of the *inverse problem* should be used.
- Several improvements can be made to condition an ill-posed problem by the incorporation of prior information on decision variables parameters. The minimum leak size or the valve status that can be detected and localised is determined by: (a) the pressure change caused by the fault, and (b) the variation caused by the errors/uncertainty in pressure measurement readings. As long as the former exceeds the latter, any leak or valve status can be detected and localised. Moreover, it is shown in this thesis that when prior information associated with the decision variables is used following search space reduction, the condition of an ill-posed problem will always improve. Still, there is no guarantee that an improvement will be good enough to convert an ill-posed problem into a well-posed one.
- The calibration performance criteria need to be revised for an improved leak detection and localisation. Incorrect conclusions can be drawn even after the calibration criteria have been met. An important aspect in the effectiveness of the model calibration is the consideration of demand fluctuation as a result of the network characteristics. Furthermore, a good model fit (measured by some objective function value) does not always solve the calibration problem. As shown in the semi real case study SR2 analysed here, it may lead to unreliable results. Therefore, traditional modelling calibration methods are too coarse to routinely provide models for reliable leakage detection. The scope for error due to system and data anomalies has been demonstrated in SR2.
- The real case study results have shown that the developed methods have a great potential not only for aiding the determination of the size and location of new leak/burst events, but also to improve the overall model quality. Through a synergy with externally-based methods, it is possible to reduce the search area and reduce the cost of finding the leak and the

volume of water lost. In practice, it should lead to a useful support tool for network operations.

6.5 Further Work Recommendations

The present research has identified a number of areas where further work can be undertaken. Firstly, it should be noted that the approaches presented in this thesis were only tested on four real-life networks with both artificially generated and real data. Additional tests on other WDNs should be considered, to establish that the results and conclusions are not network-specific. As noted in the review of the state of the art (Chapter 2), research on WDN calibration is continuously advancing with the search for the best methodology still continuing. The proposed methods provide a practical alternative that satisfied the requirement for most of the possible uses of WDN hydraulic models. This is because they can work synergistically with externally-based leak detection and localisation methods to find leaks in an effective and timely manner. However, further improvements are still possible.

6.5.1 Leak Detection and Localisation

Regarding leak detection and localisation the recommendations for further research work are as follows:

- The leak detection method has proved to correctly detect and locate a real leak. Furthermore, the analysis with synthetic data has extracted additional conclusions about detectability of leaks. A future research should assess more thoroughly the robustness and the level of leak localisation accuracy of this methodology, as well as its applicability in engineering practice. This can be accomplished through a series of desktop experiments that systematically test the method's accuracy on different case studies. The LIM can be tested on a variety of scenarios, such as for different network size and complexity, different leak size, number, and locations, as well as different number and locations of sensor devices. It is also of interest to test and validate the methodology on the

detection and localisation of multiple leaks to determine the ability to detect simultaneous events.

- There is also a need to investigate any potential applications of this methodology in near real-time so that the trade-off benefit between the leak localisation accuracy and the available data can be benchmarked. The accuracy of the LIM in detecting and localising leaks can be assessed on a number of desktop and real cases, for different scenarios on the duration and frequency interval of the recorded pressure and flow data.
- A detailed comparison of the performance of this method with previously published inverse methods should be performed. A number leak detection and localisation problems can be solved that compare this method's accuracy and computational performance against the methods developed by Wu et al. (2010) and Nasirian et al. (2013). All methods can be tested for different leak sizes, numbers and locations.
- Equally, a detailed comparison of the performance of the LIM can be performed compared with the optimisation-based methods used in the widely available commercial hydraulic modelling software such as WaterGEMS and Synergi.
- Future work could also be undertaken to determine the effect of DMAs in a cascaded system. A leak/burst event in a DMA may cause a change in pressure in other DMAs that are in the cascaded system. Therefore, the question is how will the correct DMA be identified? To better understand the effect of leak/burst events in cascaded systems, field trials should also be conducted.

6.5.2 Model Calibration

Regarding calibration of WDN hydraulic models recommendations for further research work are as follows:

- To assess the robustness of this methodology and its transferability to engineering practice after being tested on different case studies. A set of difficult, multimodal, calibration problems can be developed that test the method's accuracy based on different network situations. The CM can be

tested for different number and locations of incorrectly modelled valves, as well as different number and location of sensor devices.

- Benchmarking of the performance of this method with previously published calibration approaches should be performed. The CM can be compared with the traditional calibration approach used by practitioners (Puust & Vassiljev, 2014). In addition with approaches that used other optimisation techniques (Dini & Tabesh, 2014), or those where model calibration was combined with leak detection (Wu & Clark, 2009).
- To validate the effect of a chosen calibration dataset on the outcome of the calibration methodology. This can be achieved by testing the CM on a number of real case studies, of different range of demand fluctuation, where different datasets can be chosen to calibrate the WDN hydraulic model.
- To consider formulating the calibration problem as a two objective problem. The two objectives would minimize the weighted sum of squared differences considering the observed and validation data, respectively. This will help with understanding the possibility of overfitting.

6.5.3 Optimisation Technique and Data Effect

To further develop this work on a number of different research areas associated with the optimisation techniques could be explored on the outcome of calibration and leak localisation:

- It would be of interest to compare the benefits of optimal instrument location on the outcome of both calibration and leak localisation against traditional sensor placement methods on a real case study. For example, one or more engineered events can be introduced in a real WDN, such as a leak and/or an unknown closed/open valve.
- Because the Genetic Algorithm approach is a stochastic-search technique, the solutions obtained from different random seed values may be different. Further work is required to implement the proposed methods using other evolutionary and non-evolutionary optimisation techniques and compare them against the Genetic Algorithm optimisation.

- In order to improve the exploration and the exploitation of the solution space, a hybrid optimisation could be developed. The main idea would be the use of a genetic algorithm for global search, followed by a local search, such as a gradient method. In addition, a faster optimisation method compared to a Genetic Algorithm can be tested as an alternative for carrying out the third step of the search space reduction method.

In conclusion, it is hoped that this thesis has provided a further insight into the challenges and complexities of the *inverse problems* in Water Distribution Network Modelling. Furthermore, it may have contributed in a small way to the better understanding and solution of these problems, thereby being worthwhile on both academic and practical grounds.

APPENDIX A Leakage Inspection Tool

Appendix A provides insight on the developed Leakage Inspection Tool. It is aimed to be used as a guide for detecting and localising leaks using the tool. The presented information is divided into three main sections A1, A2 and A3. An introduction to the Leakage Inspection Tool is provided in section A1. Then, section A2 outlines the prerequisites for using the tool, before any analysis is undertaken. Finally, in section A3 the step-by-step user instructions are presented for carrying out the leak detection and localisation process.

A.1 Tool Description

The Leakage Inspection Tool is the computer-based system that was developed as an output from the proposed Leakage Inspection Method. It encapsulates the two stages of the decision-support framework for solving the *inverse problem* of leak detection and localisation, whereby the problem is reduced before being solved. Therefore, the number of candidate leak locations and the corresponding range of possible flow values is minimized before searching for any leak(s). The tool can be used for the detection and localisation of new, or existing leaks in District Metered Area (DMA) networks. This is done in two stages:

1. The Search Space Reduction Stage (SSR) for reducing the *inverse problem* size;
2. An Leak Detection and Localisation (LDL) Stage for finding leaks.

Both stages of the Leakage Inspection Tool are carried out automatically, therefore, eliminating any manual effort by the user in adjusting any hydraulic model parameter. The tool was coded using MATLAB programming language. A simulation-optimisation framework comprises it's main "engine", whereby the EPANET Programmer's Toolkit (Rossman, 2000) is used for running a hydraulic simulation and a Genetic Algorithm technique for generating and evaluating solutions. The quality of a generated solution is evaluated through the simultaneous comparison of head and flow measurements against the simulated values from the hydraulic model. Eventually, the output from this tool is a highlighted section or sections of the Water Distribution Network (WDN) where a

likely leak event has been identified, along with a proposed leak size. Consequently, the outcome can be used as support for reducing the search effort in finding the exact leak location(s) in the field with the use of externally-based methods (see Section 2.5.3).

A.2 Prerequisites and Inputs - Outputs

The prerequisites for using the Leakage Inspection Tool divide into (a) Hydraulic Model Prerequisites and (b) Data Prerequisites:

Hydraulic Model Prerequisites

1. The hydraulic model of the analysed water system should accurately represent the boundaries of the WDN, along with the number and locations of the all components.
2. Independent to whether the user would like to identify new, or existing leaks, the hydraulic model of the analysed WDN must be calibrated according to the acceptable standards (Ormsbee & Lingireddy, 1997) prior to any leakage detection. More specifically, the hydraulic model should accurately simulate the hydraulic conditions of the system before the leak new, or existing leak has happened, in order to allow its detection after it has happened.

Data Prerequisites

1. Flow measurements from at least all the inlet and outlet meters.
2. Head measurements at an indicative minimum spatial resolution of at least one sensor per 200 properties in a DMA. Therefore, for each location measurements for pressure and elevation are required.
3. The flow and head measurements must be collected at a time interval of 15 minutes and a duration of at least 24 hours to allow a full Extended Period Simulation analysis and comparison of the hydraulic model outputs.

The necessary inputs for using the Leak Inspection Tool are:

1. An EPANET version of the hydraulic model of the analysed WDN.
2. A Microsoft Excel Spreadsheet of all the recorded flow and head measurements for each monitored location. The data should be presented

in a matrix format, whereby the recorded value for each location is shown for each 15-minute time step and for a duration of at least 24 hours.

A.3 End User and Process Time Requirements

The process for using the Leakage Inspection Tool requires the end user to go through six main steps in completing a full analysis for detecting and localising leaks in a WDN. These steps are:

1. Data Preparation (Manual Step);
2. Load Inputs (Automated Step);
3. Run Step 1 and Step 2 of the SSR Stage (Automated Step);
4. Prepare and Run the Step 3 of the SSR Stage (Automated Step);
5. Prepare and Run the LDL stage (Automated Step);
6. View and Save the results (Automated Step).

1. Data Preparation

The first step in using the Leakage Inspection Tool is to gather all the flow and head measurements in a Microsoft Excel Spreadsheet in a matrix format, i.e., the recorded value every time step (columns) for each monitored location (first row). This process is carried out manually by the user and the expected completion time relies on how readily available the recorded measurements are and whether any data pre-processing is required. Assuming well-behaved measurements (see Section 4.3.2) have been collected, downloaded and are already available on the user's computer, the Data Preparation process takes only five to ten minutes. This is because the data from each sensor only need to be copied and pasted into the spreadsheet that will be loaded into the tool. On the other hand, if the monitored DMA is densely monitored, and/or the data need to be pre-processed to reach a well-behaved status, the process may require additional time.

2. Load Inputs

The second step in this analysis is to load the inputs into the Leakage Inspection Tool. The EPANET version of the hydraulic model of the analysed WDN and the Excel spreadsheet of the recorded measurements used for leak detection and localisation are inserted automatically into the tool, by pressing the related button.

3. Run Step 1 and Step 2 of the SSR Stage

During the third step the Inverse Problem Simplification and Parameter Sensitivity Analysis are carried out automatically by pressing the relevant button. This step only uses the loaded hydraulic model and not the observed data. Before running the process, the user needs to specify starting leak size for initiating the Parameter Sensitivity Analysis as a percentage value of the average inlet flow. This step can take between 15-45 minutes depending on the number of hydraulic simulations required for the Parameter Sensitivity Analysis (CM_{MDNL} in section 5.4.1.1) and the size of the hydraulic model (i.e., the complexity of the hydraulic simulation analysis).

4. Prepare and Run Step 3 of the SSR Stage

In the fourth step, the user needs to define the number and parameters of the optimisation analyses that will be carried out in Part I and Part II of the Search Space Optimisation step. The analysis is then, undertaken automatically. More specifically, for Part I the user must specify:

(a) the number of optimisation analyses (Z in equation 5.4);
(b) the emitter coefficient range of adjustment (d_z in section 3.5.2.2.) for each optimisation analysis z.
(c) the solution population size (PS in equation 5.4);
(d) the number of generations (G in equation 5.4);
(e) the solution crossover fraction;
(f) the solution mutation fraction.

For Part II the user must specify the number of optimisation analyses Z carried out for each tested leak scenario (n in section 3.5.2.3). In this research the corresponding Z values in the fourth step were specified as (a) Part I: $Z = 5$ and, (b) Part II $Z = 3$, therefore the analyses were completed within 30 – 45 minutes.

5. Prepare and Run the LDL Stage

As in the previous step, during the fifth step, the user must specify the following parameters that are necessary to carry out the LDL stage:

(a) the total number of times that the optimisation analyses will be carried out;
(b) the solution population size;
(c) the number of generations;

(d) the solution crossover fraction;

(e) the solution mutation fraction.

The LDL stage is undertaken automatically by pressing the relevant button, and total time required to complete it relies on the specified numbers for a, b and c. It is expected that the LDL stage will require at least one hour.

6. View and Save the results

Eventually, after all previous steps have been completed the user can view and save the results of all the optimisation analyses that have been carried out during the fifth step.

APPENDIX B Calibration Tool

Appendix B presents the equivalent information required to use the Calibration Tool. This tool can be used to optimise a DMA hydraulic model so that it accurately simulates the average day hydraulic conditions in the WDN and hence can be used with more confidence for leak detection and localisation. It is expected that using the instructions in Appendix B, the end user will be able to carry out a calibration analysis of a DMA hydraulic model using this tool.

B.1 Tool Description

The Calibration Tool is the computer-based system that was developed as an output from the proposed Calibration Method. It determines the state of internal pipe roughness values and detects the status of any throttle valve with uncertain position. It also identifies the setting/speed of any pressure reducing valve/variable speed pump and the multiplier coefficient value of any demand pattern in a DMA network. As in the Leak Inspection Tool, the two stage decision-support framework for solving the *inverse problem* of hydraulic model calibration is also encapsulated in the Calibration tool. Therefore, the number of candidate closed valve locations, as well as the number of candidate pipe roughness calibration groups and their corresponding range of roughness values is minimized before searching for the calibration analysis. This is done in four stages:

1. A Data Pre-Processing (DPP) Stage for generating the calibration dataset;
2. A Profile Calibration (PC) Stage for macro-level calibration of the WDN model.
3. A Search Space Reduction Stage (SSR) for reducing the calibration problem size;
4. A Component Calibration (CC) Stage for micro-level calibration.

The first stage of the Calibration Tool is carried out manually and can be the most laborious process required by the end user. In contrast Stage 2, 3 and 4 are carried out automatically, eliminating the manual effort by the end user in adjusting any hydraulic model parameter. The MATLAB programming language was also used in creating this tool, which linked the EPANET Programmer's

Toolkit (Rossman, 2000) with a Genetic Algorithm technique. Similarly to the Leakage Inspection Tool, a solution is evaluated through the simultaneous comparison of head and flow measurements against the simulated values from the hydraulic model. The output from this tool is a calibrated hydraulic model of the analysed WDN. Such model simulates the average hydraulic conditions of the system within the acceptable standards (Ormsbee & Lingireddy, 1997), thus, meeting the prerequisite hydraulic model quality so it can be used for leak detection and localisation purposes by the Leakage Inspection Tool.

B.2 Prerequisites and Inputs - Outputs

The prerequisites for using the Calibration Tool, are similar the Leakage Inspection Tool's prerequisites with the only difference being that there is no need for a calibrated hydraulic model.

The necessary inputs for using the Leak Inspection Tool are:

1. An EPANET version of the hydraulic model of the analysed WDN.
2. A Microsoft Excel Spreadsheet of the calibration dataset comprised of flow and head measurements for each monitored location. The calibration dataset should be presented in a matrix format, whereby the recorded value for each location is shown for each 15-minute time step and for a total duration of 24 hours.

B.3 End User and Process Time Requirements

The process for using the Calibration Tool also requires the end user to go through six main steps in completing the procedure of producing a model that simulates the WDN hydraulic conditions of an average day. These steps are:

1. Raw Data Preparation (Manual Step)
2. Calibration Dataset Preparation (Manual Step);
3. Load Inputs (Automated Step);
4. Prepare and Run the DPC Stage (Automated Step);
5. Prepare and Run the PPC Stage (Automated Step);
6. Prepare and Run the SSR Stage (Automated Step);
7. Prepare and Run the CC stage (Automated Step);
8. View and Save the results (Automated Step).

1. Raw Data Preparation

The first step in using the Calibration Tool is to gather the dataset of all the raw measurements (e.g., for five weekdays) for flow and head into a Microsoft Excel spreadsheet in a matrix format similar to the Leakage Inspection Tool. The same manual process as for the Leakage Inspection Tool is carried out therefore the expected workload is similar.

2. Calibration Dataset Preparation

During the second step the calibration dataset is prepared by pre-processing the raw measurements. Any corrupt, or missing values in the raw dataset is diagnosed and treated so that a well-behaved dataset is available (see sections 4.3.2 and 4.3.3). Applying the two-step procedure of the DPP stage on the raw measurements will then, create a 24-hour de-noised dataset that composes the calibration dataset. This DPP stage is carried out manually and the end user workload depends on the amount of pre-processing that must be carried out to ultimately create the 24-hour calibration dataset. Assuming well-behaved raw measurements (see Section 4.3.2) are already available, the Data Pre-Processing stage process can be completed within 15-20 minutes. The data from each sensor for each day is averaged to create the Average Day Profile Dataset (see section 4.5.2.3). Then, the De-Noised Dataset (see section 4.5.2.4), which is then used in the following steps of the Calibration Tool is generated after applying a smoothing technique. However, if the monitored DMA is densely monitored, and/or the data need to be pre-processed to reach a well-behaved status, the process may require additional time.

3. Load Hydraulic Model

The third step in this procedure is to load the Excel spreadsheet of the calibration measurements, prepared during the previous step, into the Calibration tool, along with EPANET version of the hydraulic model of the analysed WDN. This is carried out automatically, by pressing the relevant button.

4. Prepare and Run DPC Stage

In the fourth step, the user needs to define the demand patterns that will be calibrated along with the optimisation analysis parameters. The analysis is then

undertaken automatically for each pattern multiplier. More specifically, the following need to be specified:

a. the DMA Demand Profile Names considered for calibration;
b. the solution population size at each time step;
c. the number of generations at each time step;
d. the solution crossover fraction;
e. the solution mutation fraction.

This analysis requires a large computational burden, as an optimisation analysis is carried out for each time step. Therefore, it can take between 2 – 5 hours depending on the specified population size and the number of generations for each time step, as well as the size of the hydraulic model (i.e., the complexity of the hydraulic simulation analysis).

5. Prepare and Run PPC Stage

The fifth step in using the Calibration Tool is very similar to the previous step. The end user must specify the equivalent optimisation analysis parameters. The difference lies on the fact that in the PPC Stage the names of all pressure reducing valves and variable speed pumps that will be calibrated need to be defined. It is expected that a similar computational time is required for completing this step.

6. Prepare and Run the SSR Stage

The Inverse Problem Simplification and Parameter Sensitivity Analysis are carried out automatically by pressing the relevant button. The Parameter Sensitivity Analysis can be completed within 5 – 15 minutes for either the valve or, pipe components. The required computational burden in the Parameter Sensitivity Analysis depends on the number of tested parameters (CM_{DVC} and CM_{DPRC} in section 5.4.1.1.) and the size of the hydraulic model (i.e., the complexity of the hydraulic simulation analysis). The Search Space Optimisation step that is carried out after the Parameter Sensitivity Analysis is also automated. At this step the user only needs to define the parameters of the optimisation analyses that will be carried out:

(a) the solution population size;
(b) the number of generations;

(c) the solution crossover fraction;

(d) the solution mutation fraction.

Search Space Optimisation may require between 30 – 60 minutes for valve components and 10 – 30 minutes for pipe components. This depends on the specified parameters for a, b and c, as well as the number of candidate valves following the Parameter Sensitivity Analysis.

7. Prepare and Run the CC Stage

As in the previous step, during the CC stage, the user must specify the optimisation analysis parameters. In addition, the number of optimisation analyses that will be undertaken. The CC stage is, then, performed automatically, and the required computational time is at least one hour.

8. View and Save the results

Eventually, after all previous steps have been completed the user can view and save the results of all the optimisation analyses that have been carried out during the seventh step.

Bibliography

ACAPS, 2016. *Data Cleaning - Dealing with messy data*, Available at: https://www.acaps.org/sites/acaps/files/resources/files/acaps_technical_brief_data_cleaning_april_2016_0.pdf.

ADEC, 1999. *Technical review of leak detection technologies - vol.1 - Crude Oil Transmission Pipelines*, Alaska.

Aksela, K., Aksela, M. & Vahala, R., 2009. Leakage detection in a real distribution network using a SOM. *Urban Water Journal*, 6(4), pp.279–289.

Almandoz, J. et al., 2005. Leakage Assessment through Water Distribution Network Simulation. *Journal of Water Resources Planning and Management*, 131(6), pp.458–466. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%282005%29131%3A6%28458%29.

Alvisi, S. & Franchini, M., 2010. Pipe roughness calibration in water distribution systems using grey numbers. *Journal of Hydroinformatics*, 12(4), p.424. Available at: http://jh.iwaponline.com/cgi/doi/10.2166/hydro.2010.089.

Aral, M.M., Guan, J. & Maslia, M.L., 2010. Optimal Design of Sensor Placement in Water Distribution Networks. *Journal of Water Resources Planning and Management*, 136(1), pp.5–18. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000001.

Armon, A. et al., 2011. Algorithmic network monitoring for a modern water utility: A case study in Jerusalem. *Water Science and Technology*, 63(2), pp.233–239.

AWWA, 1999. *Calibration Guidelines for Water Distribution System Modeling*, Available at: http://www.awwa.org/unitdocs/592/calibrate.pdf.

Bäck, T., Hammel, U. & Schwefel, H.P., 1997. Evolutionary computation: Comments on the history and current state. *IEEE Transactions on Evolutionary Computation*, 1(May), pp.3–17.

Barandouzi, M.A. et al., 2012. Probabilistic Mapping of Water Leakage Characterizations Using a Bayesian Approach. In *World Environmental And Water Resources Congress*. New Mexico, USA: American Society of Civil

Engineers.

Batini, C. & Scannapieco, M., 2006. *Data Quality* 1st ed., Springer-Verlag Berlin Heidelberg.

Baumeister, J., 1987. *Stable Solution of Inverse Problems* 1st ed., Vieweg+Teubner Verlag.

Beiranvand, V., Hare, W. & Lucet, Y., 2017. Best practices for comparing optimization algorithms. *Optimization and Engineering*, 18(4), pp.815–848.

Bekey, G. & Kogan, B.J., 2003. *Modeling and Simulation: Theory and Practice*, Springer US.

van den Berg, C., 2014. *The Drivers of Non-Revenue Water How Effective Are Non-Revenue Water Reduction Programs?*, Available at: http://www-wds.worldbank.org/external/default/WDSContentServer/IW3P/IB/2014/08/05/000158349_20140805090718/Rendered/PDF/WPS6997.pdf.

Berglund, A. et al., 2017. Successive Linear Approximation Methods for Leak Detection in Water Distribution Systems. *Journal of Water Resources Planning and Management*, 143(8), p.04017042. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000784.

Bertero, M., Mol, C.D. & Viano, G.A., 1980. The Stability of Inverse Problems. In H. P. Baltes, ed. *Inverse Scattering Problems in Optics*. Berlin: Springer-Verlag Berlin Heidelberg.

Bhave, P.R., 1988. Calibrating Water Distribution Network Models. *Journal of Environmental Engineering*, 114(1), pp.120–136.

Bilicz, S., Lambert, M. & Gyimóthy, S., 2010. Inverse problem characterization using an adaptive database. In *20th URSI International Symposium on Electromagnetic Theory*. Berlin: IEEE, pp. 613–616.

Black, P., 1992. A Review of Pipeline Leak Detection Technology. In B. Coulbeck & E. P. Evans, eds. *Fluid Mechanics and its Applications*. Dordrecht: Springer.

Blesa, J., Nejjari, F. & Sarrate, R., 2016. Robust sensor placement for leak location: analysis and design. *Journal of Hydroinformatics*, 18(1), pp.136–148. Available at: http://jh.iwaponline.com/lookup/doi/10.2166/hydro.2015.021.

Boulos, P.F. & Ormsbee, L.E., 1991. Explicit network calibration for multiple loading conditions. *Civil Engineering Systems*, 8(3), pp.153–160.

Boulos, P.F. & Wood, D.J., 1990. Explicit calculation of pipe network parameters. *Journal of Hydraulic Engineering*, 116(11), pp.1329–1344.

Van Den Broeck, J. et al., 2005. Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Medicine*, 2(10), pp.0966–0970.

Burrows, R., 1999. Demand for network model building.

Burrows, R. et al., 2003. Introduction of a fully dynamic representation of leakage into network modelling studies using Epanet. In C. Maksimovic, D. Butler, & F. A. Memon, eds. *Advances in Water Supply Management Proceedings of the CCWI '03 Conference*. London.

Camponogara, E. & Talukdar, N., 1997. *A Genetic Algorithm for Constrained and Multiobjective Optimisation*, Helsinki, Finland.

Candelieri, A., Contib, D. & Archetti, F., 2014. A graph based analysis of leak localization in urban water networks. *Procedia Engineering*, 70, pp.228–237. Available at: http://dx.doi.org/10.1016/j.proeng.2014.02.026.

Caputo, A.C. & Pelagagge, P.M., 2002. An inverse approach for piping networks monitoring. *Journal of Loss Prevention in the Process Industries*, 15(6), pp.497–505.

Caputo, A.C. & Pelagagge, P.M., 2003. Using Neural Networks to Monitor Piping Systems. *Process Safety Progress*, 22(2), pp.119–127.

Carrera, J. & Neuman, S.P., 1986. Estimation of Aquifer Parameters Under Transient and Steady State Conditions' Uniqueness, Stability, and Solution Algorithms. *Water Resources Research*, 22(2), pp.211–227.

Casillas, M. et al., 2015. Leak Signature Space: An Original Representation for Robust Leak Location in Water Distribution Networks. *Water*, 7(3), pp.1129–1148. Available at: http://www.mdpi.com/2073-4441/7/3/1129/.

Casillas Ponce, M.V., Castañón, L.E.G. & Cayuela, V.P., 2014. Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks. *Journal of Hydroinformatics*, 16(3), pp.649–670.

Casillas Ponce, M.V., Castañón, L.E.G. & Cayuela, V.P., 2012. Extended-horizon analysis of pressure sensitivities for leak detection in water distribution networks. In *Symposium on Fault Detection*. Mexico City: International Federation of Automatic Control, pp. 570–575.

Cesario, A.L. et al., 1996. New Perspectives on Calibration of Treated Water Distribution System Models. In *Proceedings of the AWWA Annual*

Conference. Toronto, Canada: American Water Works Association.

Chen, K. & Brdys, M.A., 1995. Set Membership Estimation of State and Parameters in Quantity Models of Water Supply and Distribution Systems. In *Proceedings of IFAC Symposium on Large Scale Systems*. London: Elsevier, pp. 543–548. Available at: http://dx.doi.org/10.1016/S1474-6670(17)51575-X.

Cheng, W. & He, Z., 2011. Calibration of Nodal Demand in Water Distribution Systems. *Journal of Water Resources Planning and Management*, 137(1), pp.31–40.

Clark, A., 2012. Increasing Efficiency with Permanent Leakage Monitoring. In *Water Loss Reduction*. 7th IWA Specialist Conference, pp. 26–29.

Cleveland, W.S., 1979. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association*, 74(368), pp.829–836.

Colombo, A.F., Lee, P. & Karney, B.W., 2009. A selective literature review of transient-based leak detection methods. *Journal of Hydro-Environment Research*, 2(4), pp.212–227. Available at: http://dx.doi.org/10.1016/j.jher.2009.02.003.

Costanzo, F. et al., 2014. Model calibration as a tool for leakage identification in WDS: A real case study. *Procedia Engineering*, 89, pp.672–678.

Covas, D.I.C., 2003. *Inverse transient analysis for leak detection and calibration of water pipe systems modelling special dynamics effects*. University of London.

Cugueró-Escofet, P. et al., 2015. Assessment of a leak localization algorithm in water networks under demand uncertainty. *International Federation of Automatic Control*, 48(21), pp.226–231.

Dasu, T. & Johnson, T., 2003. *Exploratory Data Mining and Data Cleaning*, New Jersey: John Wiley & Sons, Inc. Available at: http://www.tandfonline.com/doi/abs/10.1198/jasa.2006.s81.

Deb, K. et al., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), pp.182–197.

Debiasi, S. et al., 2014. Influence of hourly water consumption in model calibration for leakage detection in a WDS. *Procedia Engineering*, 70, pp.467–476. Available at: http://dx.doi.org/10.1016/j.proeng.2014.02.052.

Dini, M. & Tabesh, M., 2014. A New Method for Simultaneous Calibration of Demand Pattern and Hazen-Williams Coefficients in Water Distribution Systems. *Water Resources Management*, 28(7), pp.2021–2034.

Discover Water, 2018. Leaking Pipes. Available at: https://discoverwater.co.uk/leaking-pipes [Accessed November 16, 2018].

DNV-GL, 2018. Water quality analysis and hydraulic modelling - Synergi Water. Available at: https://www.dnvgl.com/services/water-quality-analysis-and-hydraulic-modelling-synergi-water-2792?utm_campaign=pipeline_synergi_water&utm_source=google&utm_medium=cpc&gclid=Cj0KCQiA3IPgBRCAARIsABb-iGJiP_RVtytzGhqDtcWIvPXpoYDrkPjqX4gZQDSVZyOeqTBa3osYkYoaAiX [Accessed November 30, 2018].

Do, N.C. et al., 2016. Calibration of Water Demand Multipliers in Water Distribution Systems Using Genetic Algorithms. *Journal of Water Resources Planning and Management*, 142(11), p.04016044. Available at: http://ascelibrary.org/doi/10.1061/(ASCE)WR.1943-5452.0000691.

Echologics, 2015. LeakFinder-ST™ Correlator. Available at: https://www.echologics.com/products/leakfinderst/ [Accessed November 26, 2018].

Ediriweera, D.D. & Marshall, I.W., 2010. Monitoring water distribution systems: Understanding and managing sensor networks. *Drinking Water Engineering and Science*, 3(2), pp.107–113. Available at: www.drink-water-eng-sci.net/3/107/2010/.

Environment Agency, 2008. *Water resources in England and Wales - current state and future pressures*, Bristol. Available at: www.environment-agency.gov.uk.

Fanner, P. V. et al., 2007. *Leakage Management Technologies*, Washington: AWWA Research Foundation, American Water Works Association, International Water Association Publishing.

Farley, B., Boxall, J.B. & Mounce, S.R., 2008. Optimal Locations of Pressure Meters for Burst Detection. In *Water Distribution Systems Analysis 2008*. Kruger National Park, South Africa: American Society of Civil Engineers.

Farley, B., Mounce, S.R. & Boxall, J.B., 2013. Development and Field Validation of a Burst Localization Methodology. *Journal of Water Resources Planning*

and Management, 139(December), pp.604–613.

Farley, B., Mounce, S.R.M. & Boxall, J.B., 2012. Development and field validation of a burst localisation. *Journal of Water Resources Planning and Management*, 139(6), pp.604–613.

Farley, M. & Trow, S., 2003. *Losses in Water Distribution Networks: A Practitioner's Guide to Assessment, Monitoring and Control*, London: International Water Association Publishing.

Farmani, R., Savic, D.A. & Walters, G.A., 2005. Evolutionary multi-objective optimization in water distribution network design. *Engineering Optimization*, 37(2), pp.167–183.

Feng, J. & Zhang, H., 2006. Algorithm of Pipeline Leak Detection Based on Discrete Incremental Clustering Method. In D. S. HuangK., K. Li, & G. W. Irwin, eds. Berlin: Springer-Verlag Berlin Heidelberg, pp. 602–607.

Ferrandez-Gamot, L. et al., 2015. Leak Localization in Water Distribution Networks using Pressure Residuals and Classifiers. *International Federation of Automatic Control*, 48(21), pp.220–225. Available at: https://linkinghub.elsevier.com/retrieve/pii/S2405896315016602.

Ferreri, G., Napoli, E. & Tumbiolo, A., 1994. Calibration of roughness in water distribution networks. In D. S. Miller, ed. *Proceedings of International Conference on Water Pipeline Systems*. Edinburgh, pp. 379–396.

Field, D.B. & Ratcliffe, B., 1978. *Location of leaks in pressurised pipelines using sulphur hexafluoride as a tracer*,

Fuchs, H. V. & Riehle, R., 1991. Ten years of experience with leak detection by acoustic signal analysis. *Applied Acoustics*, 33(1), pp.1–19.

Furness, R.A. & Reet, J.D., 1998. *Pipe line leak detection techniques* Pipe Line. E. W. McAllister, ed., Houston, Texas: Gulf Publishing Company.

Gamboa-Medina, M.M. & Reis, L.F.R., 2017. Sampling Design for Leak Detection in Water Distribution Networks. *Procedia Engineering*, 186, pp.460–469. Available at: http://dx.doi.org/10.1016/j.proeng.2017.03.255.

General Electric, 2018. DPI 611 Hand-held Pressure Calibrator. *GE Measurement & Control*. Available at: https://www.gemeasurement.com/sites/gemc.dev/files/920-651b-hr.pdf [Accessed November 16, 2018].

Germanopoulos, G., 1985. A technical note on the inclusion of pressure

dependent demand and leakage terms in water supply network models. *Civil Engineering Systems*, 2(3), pp.171–179.

Giustolisi, O., Savic, D. & Kapelan, Z., 2008. Pressure-Driven Demand and Leakage Simulation for Water Distribution Networks. *Journal of Hydraulic Engineering*, 134(5), pp.626–635. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9429%282008%29134%3A5%28626%29.

Gleick, P.H., 1995. Human population and water: to the limits in the 21st century. *American Association for the Advancement of Science*. Available at: http://www.aaas.org/international/ehn/fisheries/gleick.htm [Accessed November 16, 2018].

Goldberg, D.E., 1989. *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison-Wesley Publishing Co.

Goodwin, S.J., 1980. *The Results of the Experimental Programme on Leakage and Leakage Control*, Water Research Center Environmental Protection.

Goulet, J.A., Coutu, S. & Smith, I.F.C., 2013. Model falsification diagnosis and sensor placement for leak detection in pressurized pipe networks. *Advanced Engineering Informatics*, 27(2), pp.261–269. Available at: http://dx.doi.org/10.1016/j.aei.2013.01.001.

Greco, M. & Del Giudice, G., 1999. New Approach to Water Distribution Network Calibration. *Journal of Water Resources Planning and Management*, 125(8), pp.849–854.

Groetsch, C.W., 1993. *Inverse Problems in the Mathematical Sciences* 1st ed., Wiesbaden: Vieweg+Teubner Verlag.

Guttermann, 2014. AQUASCAN 610 – LEAK NOISE CORRELATOR. Available at: https://en.gutermann-water.com/product/aquascan-610-leak-noise-correlator/ [Accessed November 26, 2018].

Hadamard, J., 1923. Lectures on the Cauchy Problem in Linear Partial Differential Equations.

Halaczkiewicz, T. & Klima, P., 2018. Demystifying the pressure gauge spec sheet. *Flow Control*. Available at: https://www.flowcontrolnetwork.com/demystifying-the-pressure-gauge-spec-sheet/ [Accessed November 16, 2018].

Hamilton, S. & Charalambous, B., 2013. *Leak Detection: Technology and*

Implementation 1st ed., London: International Water Association Publishing. Available at: https://books.google.com/books?id=lYnHCgAAQBAJ&pgis=1.

Hamilton, S., Krywyj, D. & Jones, C., 2012. The Problem of Leakage Detection on Large Diameter Mains. In *Pipelines Conference*. American Society of Civil Engineers.

Hartley, D., 2009. Acoustics Paper. In *Water Loss Reduction*. Cape Town: 5th IWA Specialist Conference.

Hassan, R., Cohanim, B. & de Weck, O., 2005. A Comparison of Particle Swarm Optimization and the Genetic Algorithm. In *Structural Dynamics & Material Conference*. pp. 274–283.

Hill, M.C., 1998. *Methods and Guidelines for*, Denver, Colorado.

Holland, J.H., 1975. *Adaption in Natural and Artificial Systems*, MIT Press.

Hoos, H.H. & Stützle, T., 2004. *Stochastic Local Search: Foundations and Applications* 1st ed., Burlington: Morgan Kaufmann.

Hugenschmidt, J. & Kalogeropoulos, A., 2009. The inspection of retaining walls using GPR. *Journal of Applied Geophysics*, 67(4), pp.335–344. Available at: http://dx.doi.org/10.1016/j.jappgeo.2008.09.001.

Hunaidi, O., 2000. *Detecting Leaks in Water-Distribution Pipes*, Ottawa.

Hunaidi, O. & Chu, W.T., 1999. Acoustical characteristics of leak signals in plastic water distribution pipes. *Applied Acoustics*, 58(3), pp.235–254.

Hutton, C.J. et al., 2014. Dealing with Uncertainty in Water Distribution System Models: A Framework for Real-Time Modeling and Data Assimilation. *Journal of Water Resources Planning and Management*, 140(2), pp.169–183. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000325.

Hydraulic Research, 1983. Tables for the Hydraulic Design of Pipes and Sewers.

Insurance Times, 2013. Insurers face £4m payout for south London flood. Available at: https://www.insurancetimes.co.uk/insurers-face-4m-payout-for-south-london-690 [Accessed November 16, 2018].

Jacoby, W.G., 2000. Loess: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, 19(4), pp.577–613.

De Jong, K.A., 1975. *Analysis of the behavior of a class of genetic adaptive*

systems. The University of Michigan.

Kabanikhin, S.I., 2008. Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-Posed Problems*, 16(4), pp.317–357.

Kapelan, Z., 2002. *Calibration of Water Distribution System Hydraulic Models*. University of Exeter.

Kapelan, Z., Savic, D. & Walters, G., 2004. Incorporation of prior information on parameters in inverse transient analysis for leak detection and roughness calibration. *Urban Water Journal*, 1(2), pp.129–143. Available at: http://www.tandfonline.com/doi/abs/10.1080/15730620412331290029.

Kapelan, Z.S., Savic, D.A. & Walters, G.A., 2003. A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks. *Journal of Hydraulic Research*, 41(5), pp.481–492.

Kapelan, Z.S., Savic, D.A. & Walters, G.A., 2007. Calibration of Water Distribution Hydraulic Models Using a Bayesian-Type Procedure. *Journal of Hydraulic Engineering*, 133(8), pp.927–936. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9429%282007%29133%3A8%28927%29.

Karimova, F., 2016. Intelligent Water Systems: A Smart Start. Available at: https://www.wef.org/globalassets/assets-wef/3---resources/online-education/webcasts/presentation-handouts/iws-webcast---presentation-handouts.pdf.

Kennedy, J. & Eberhart, R., 1995. Particle Swarm Optimization. In *Proceedings of the IEEE International Conference on Neural Networks*. Perth, WA, Australia: IEEE, pp. 1942–1948.

Kingdom, B., Liemberger, R. & Marin, P., 2006. The challenge of reducing non-revenue water (NRW) in developing countries - How the private sector can help : A look at performance-based service contracting. *Water Supply and Sanitation Sector Board Discussion Paper Series*, (8), pp.1–52. Available at: http://documents.worldbank.org/curated/en/2006/12/7531078/challenge-reducing-non-revenue-water-nrw-developing-countries-private-sector-can-help-look-performance-based-service-contracting.

Kirkpatrick;, S., Gelatt;, C.D. & Vecchi, M.P., 1983. Optimization of Simulated Annealing. *Science*, 220(4598), pp.671–680. Available at: http://www.mendeley.com/research/no-title-avail/.

Kool, J.B., Parker, J.C. & van Genuchten, M.T., 1987. Parameter estimation for unsaturated flow and transport models - A review. *Journal of Hydrology*, 91(3–4), pp.255–293.

Koppel, T. & Vassiljev, A., 2009. Advances in Engineering Software Calibration of a model of an operational water distribution system containing pipes of different age. *Advances in Engineering Software*, 40(8), pp.659–664. Available at: http://dx.doi.org/10.1016/j.advengsoft.2008.11.015.

Koppel, T. & Vassiljev, A., 2012. Use of modelling error dynamics for the calibration of water distribution systems. *Advances in Engineering Software*, 45(1), pp.188–196. Available at: http://dx.doi.org/10.1016/j.advengsoft.2011.09.024.

Kun, D. et al., 2018. Direct Inversion Algorithm for Pipe Resistance Coefficient Calibration of Water Distribution Systems. *Journal of Water Resources Planning and Management*, 144(7). Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000948.

Kun, D. et al., 2015. Inversion Model of Water Distribution Systems for Nodal Demand Calibration. *Journal of Water Resources Planning and Management*, 141(9), p.04015002. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000506.

Lambert, A., 2001. What Do We Know About Pressure : Leakage Relationships in Distribution Systems ? In *System Approach to Leakage Control and Water Distribution Systems Management*. Brno, Czech Republic: International Water Association Conference.

Lambert, A., Fantozzi, M. & Shepherd, M., 2017. FAVAD Pressure & Leakage: How does pressure influence N1? *IWA Water Efficient Conference*. Available at: http://www.leakssuite.com/wp-content/uploads/2017/07/FAVAD-Pressure-Leakage-Lambert-et-al-PPT-14th-July.pdf [Accessed November 16, 2018].

Lambert, A.O., 2002. Water Losses Management and Techniques. *Water Science and Technology: Water Supply*, 2(4), pp.1–20.

Lambert, A.O. & Lalonde, A., 2005. Using practical predictions of Economic Intervention Frequency to calculate Short-run Economic Leakage Level , with or without Pressure Management. In *Leakage*. pp. 1–12.

Lansey, K.E. & Basnet, C., 1991. Parameter estimation for water distribution

networks. *Journal of Water Resources Planning and Management*, 117(1), pp.126–144.

Lewis, R.M. et al., 2000. Direct Search Methods : Then and Now Operated by Universities Space Research Association. *Journal of Computational and Applied Mathematics*, 124, pp.191–207.

Li, R. et al., 2015. A review of methods for burst/leakage detection and location in water distribution systems. *Water Science and Technology: Water Supply*, 15(3), pp.429–441.

Liggett, J.A. & Chen, L.-C., 1995. Inverse Transient Analysis in Pipe Networks. *Journal of Hydraulic Engineering*, 120(8), pp.934–955.

Loader, C., 2004. *Smoothing: Local Regression Techniques*, Humboldt. Available at: https://www.econstor.eu/bitstream/10419/22186/1/12_cl.pdf.

Mahmoud, H.A., Savić, D. & Kapelan, Z., 2017. New Pressure-Driven Approach for Modeling Water Distribution Networks. *Journal of Water Resources Planning and Management*, 143(8), p.04017031. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000781.

Mala-Jetmarova, H., Barton, A. & Bagirov, A., 2015. A history of Water distribution systems and their optimisation. *Water Science and Technology: Water Supply*, 15(2), pp.224–235.

Mamo, T.G., Juran, I. & Shahrour, I., 2014. Virtual DMA Municipal Water Supply Pipeline Leak Detection and Classification Using Advance Pattern Recognizer Multi-Class SVM. *Journal of Pattern Recognition Research*, 9(1), pp.25–42. Available at: http://www.jprr.org/index.php/jprr/article/viewFile/548/202%5Cnhttp://www.jprr.org/index.php/jprr/article/view/548.

Marques, J., Cunha, M. & Savić, D., 2015. A multicriteria approach for a phased design of water distribution networks. In *Procedia Engineering*. 13th Computer Control for Water Industry Conference, CCWI 2015, pp. 1231–1240.

Mashford, J. et al., 2009. An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. In *Network and System Security*. IEEE, pp. 1–6.

Menke, W., 2012. *Geophysical Data Analysis: Discrete Inverse Theory* 3rd ed., Academic Press.

Mergelas, B. & Henrich, G., 2005. Leak locating method for pre-commissioned transmission pipelines : North American case studies. In *Leakage*. pp. 1–7.

Meseguer, J. et al., 2014. A decision support system for on-line leakage localization. *Environmental Modelling and Software*, 60, pp.331–345.

Metropolis, N. et al., 1953. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), pp.1087–1092.

Michalewicz, Z. & Fogel, D.B., 2004. *How to Solve It: Modern Heuristics* 2nd ed., New York: Springer-Verlag Berlin Heidelberg.

Mirats-Tur, J.M. et al., 2014. Leak detection and localization using models: Field results. *Procedia Engineering*, 70, pp.1157–1165. Available at: http://dx.doi.org/10.1016/j.proeng.2014.02.128.

Moors, J. et al., 2018. Automated leak localization performance without detailed demand distribution data. *Urban Water Journal*, 15(2), pp.116–123. Available at: http://doi.org/10.1080/1573062X.2017.1414272.

Moser, G., Paal, S.G. & Smith, I.F.C., 2015. Economic inequalities in the effectiveness of a primary care intervention for depression and suicidal ideation. *Advanced Engineering Informatics*, 29, pp.714–726. Available at: http://dx.doi.org/10.1016/j.aei.2015.07.003.

Moser, G., Paal, S.G. & Smith, I.F.C., 2018. Leak Detection of Water Supply Networks Using Error-Domain Model Falsification. *Journal of Computing in Civil Engineering*, 32(2), pp.1–18.

Mounce, S.R. & Boxall, J.B., 2010. Implementation of an on-line artificial intelligence district meter area flow meter data analysis system for abnormality detection: A case study. *Water Science and Technology: Water Supply*, 10(3), pp.437–444.

Mounce, S.R., Mounce, R.B. & Boxall, J.B., 2011. Identifying Sampling Interval for Event Detection in Water Distribution Networks. *Journal of Water Resources Planning and Management*, 138(2), pp.187–191. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000170.

Mutikanga, H.E., Sharma, S.K. & Vairavamoorthy, K., 2013. Methods and Tools for Managing Losses in Water Distribution Systems. *Journal of Water Resources Planning and Management*, 139(2), pp.166–174. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000245.

Di Nardo, A. et al., 2015. A genetic algorithm for demand pattern and leakage

estimation in a water distribution network. *Journal of Water Supply: Research and Technology - AQUA*, 64(1), pp.35–46.

Nasirian, A., Maghrebi, M.F. & Yazdani, S., 2013. Leakage Detection in Water Distribution Network Based on a New Heuristic Genetic Algorithm Model. *Journal of Water Resource and Protection*, 05(03), pp.294–303. Available at: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/jwarp.2013.53030.

NEOS, 1996. Server for optimization. Available at: https://neos-server.org/neos/ [Accessed November 18, 2018].

Nyarko, E.K., Cupec, R. & Filko, D., 2014. A Comparison of Several Heuristic Algorithms for Solving High Dimensional Optimization Problems Preliminary Communication. *International Journal of Electrical and Computer Engineering Systems*, 5(1), pp.1–8.

Ofwat, 2008a. International comparison - leakage. Available at: http://www.ofwat.gov.uk/regulating/reporting/rpt_int_08leakageintro [Accessed November 16, 2018].

Ofwat, 2018. PN 20/18: Fresh thinking for managing water demand across England and Wales. Available at: https://www.ofwat.gov.uk/pn-20-18-fresh-thinking-for-managing-water-demand-across-england-and-wales/ [Accessed November 16, 2018].

Ofwat, 2008b. *Providing Best Practice Guidance on the Inclusion of Externalities in the ELL Calculation*,

Ong, A. & Rodil, M.E.H., 2012. Trunk mains leak detection in Manila's West Zone. In *Water Loss Reduction2*. 7th IWA Specialist Conference.

Ormsbee, L. & Wood, D., 1986. Explicit Pipe Network Calibration. *Journal of Water Resources Planning and Management*, 112(2), pp.166–182. Available at: http://dx.doi.org/10.1061/(ASCE)0733-9496(1986)112:2(166).

Ormsbee, L.E., 1989. Implicit network calibration. *Journal of Water Resources Planning and Management*, 115(2), pp.243–257.

Ormsbee, L.E. & Lingireddy, S., 1997. Calibrating Hydraulic Network Models. *Journal of American Water Works Association*, 89(2), pp.42–50.

Ostfeld, A. et al., 2012. Battle of the Water Calibration Networks. *Journal of Water Resources Planning and Management*, 138(5), pp.523–532. Available at: http://ascelibrary.org/doi/10.1061/41203%28425%29136.

Ostfeld, A. et al., 2008. The Battle of the Water Sensor Networks „ BWSN …: A

Design. *Journal of Water Resources Planning and Management*, 134(6), pp.556–568.

Ostfeld, A. & Salomons, E., 2004. Optimal Layout of Early Warning Detection Stations for Water Distribution Systems Security. *Journal of Water Resources Planning and Management*, 130(5), pp.377–385. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%282004%29130%3A5%28377%29.

Palau, C. V., Arregui, F.J. & Carlos, M., 2012. Burst Detection in Water Networks Using Principal Component Analysis. *Journal of Water Resources Planning and Management*, 138(1), pp.47–54. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000147.

Parker, R.L., 1977. Understanding Inverse Theory. *Annual Review of Earth and Planetary Sciences*, 5, pp.35–64.

Pearson, D. & Trow, S., 2005. Calculating the economic levels of leakage. In *Leakage*. pp. 1–16.

Pelletier, G., Mailhot, A. & Villeneuve, J.P., 2003. Modeling Water Pipe Breaks—Three Case Studies. *Journal of Water Resources Planning and Management*, 129(2), pp.115–123.

Perez, R. et al., 2011. Methodology for leakage isolation using pressure sensitivity and correlation analysis in water distribution systems. *Control Engineering Practice journal*, 19, pp.1157–1167.

Pérez, R. et al., 2014. Accuracy assessment of leak localisation method depending on available measurements. *Procedia Engineering*, 70, pp.1304–1313.

Pérez, R. et al., 2009. Leakage isolation using pressure sensitivity analysis in water distribution networks: Application to the Barcelona case study. *International Federation of Automatic Control*, 43(8), pp.578–584.

Peterson, J., 2018. Flowmeter Accuracy. *FLOMEC*. Available at: http://flomec.net/downloads/IND-1095A White Paper - Flowmeter Accuracy.pdf [Accessed November 16, 2018].

Petrov, Y.P. & Sizikov, V.V.S., 2005. *Well-posed, ill-posed, and intermediate problems with applications [electronic resource]*, Leiden: De Gruyter. Available at: http://books.google.com/books?id=J43XGdOcX8sC&pgis=1.

Pilcher, R. et al., 2007. *Leak Location and Repair Guidance Notes*,

Poulakis, Z., Valougeorgis, D. & Papadimitriou, C., 2003. Leakage detection in water pipe networks using a Bayesian probabilistic framework. *Probabilistic Engineering Mechanics*, 18(4), pp.315–327.

Pudar, R.S. & Liggett, J.A., 1992. Leaks in Pipe Networks. *Journal of Hydraulic Engineering*, 118(7), pp.1031–1046. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9429%281992%29118%3A7%281031%29.

Puust, R. et al., 2010. A review of methods for leakage management in pipe networks. *Urban Water Journal*, 7(1), pp.25–45.

Puust, R. & Vassiljev, A., 2014. Real water network comparative calibration studies considering the whole process from engineer's perspective. *Procedia Engineering*, 89, pp.702–709.

Rahal, C.M., Sterling, M.J.H. & Coulbeck, B., 1980. Parameter tuning for simulation models of water distribution networks. *Proceedings of the Institution of Civil Engineers*, 69, pp.751–762.

Ramos, H. et al., 2001. Leakage Control Policy within Operating Management Tools. In *6th International Conference on Computing and Control for the Water Industry*. Leicester, UK, pp. 61–72.

Rao, C.S. et al., 2012. Data Cleaning : A Framework for Robust Data Quality In Enterprise Data Warehouse Data Cleaning : A Framework for Robust Data Quality In Enterprise Data Warehouse. *International Journal of Computer Science and Technology*, 3(3), pp.36–41.

Reddy, P., Sridharan, K. & Rao, P., 1996. WLS Method for Parameter Estimation in Water Distribution Networks. *Journal of Water Resources Planning and Management*, 122(3), pp.157–164. Available at: http://ascelibrary.org/doi/abs/10.1061/(ASCE)0733-9496(1996)122:3(157).

Ribeiro, L. et al., 2015. Locating leaks with TrustRank algorithm support. *Water*, 7(4), pp.1378–1401.

Romano, M., Kapelan, Z. & Savic, D.A., 2009. Bayesian-based online burst detection in water distribution systems. In J. Boxall & C. Maksimovic, eds. *Integrating Water Systems*. London: Taylor and Francis Group, pp. 331–337.

Romano, M., Kapelan, Z. & Savić, D.A., 2011. Burst Detection and Location in Water Distribution Systems. In *World Environmental and Water Resources*

Congress. California: American Society of Civil Engineers.

Romano, M., Kapelan, Z. & Savić, D.A., 2014. Evolutionary Algorithm and Expectation Maximization Strategies for Improved Detection of Pipe Bursts and Other Events in Water Distribution Systems. *Journal of Water Resources Planning and Management*, 140(5), pp.572–584.

Romano, M., Kapelan, Z. & Savić, D.A., 2013. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. *Journal of Hydroinformatics*, 15(3), p.634. Available at: http://jh.iwaponline.com/cgi/doi/10.2166/hydro.2013.094.

Romeo, F. & Sangiovanni-Vincentelli, A., 1991. A theoretical framework for simulated annealing. *Algorithmica*, 6(1), pp.302–345.

Rosich, A. & Puig, V., 2013. Model-based Leakage Localization in Drinking Water Distribution Networks using Structured Residuals. In *European Control Conference*. IEEE.

Rosich, A., Puig, V. & Casillas, M. V., 2015. Leak localization in drinking water distribution networks using structured residuals Albert. *International Journal of Adaptive Control and Signal Processing*, 29, pp.991–1007.

Rossman, L., 2000. *EPANET User's Manual*, Washington. Available at: https://cfpub.epa.gov/si/si_public_record_report.cfm?Lab=NRMRL&dirEntryId=95662.

Rougier, J., 2005. Probabilistic leak detection in pipelines using the mass imbalance approach. *Journal of Hydraulic Research*, 43(5), pp.556–566.

Ruder, S., 2016. *An overview of gradient descent optimization algorithms*, Available at: http://arxiv.org/abs/1609.04747.

Russel, S. & Norvig, P., 2010. *Artificial Intelligence, A Modern Approach* 3rd ed., New Jersey: Pearson Education, Inc.

Sala, D. & Kołakowski, P., 2014. Detection of leaks in a small-scale water distribution network based on pressure data - Experimental verification. *Procedia Engineering*, 70, pp.1460–1469. Available at: http://dx.doi.org/10.1016/j.proeng.2014.02.161.

Salguero, F.J., Cobacho, R. & Pardo, M.A., 2018. Unreported leaks location using pressure and flow sensitivity in water distribution networks. *Water Science and Technology: Water Supply*, p.ws2018048. Available at: http://ws.iwaponline.com/lookup/doi/10.2166/ws.2018.048.

Santamarina, J.C. & Fratta, D., 2005. *Discrete Signal and Inverse Problems.*, Chichester: John Wiley & Sons, Inc.

Sanz, G. et al., 2016. Leak Detection and Localization through Demand Components Calibration. *Journal of Water Resources Planning and Management*, 142(2).

Sanz, G. & Pérez, R., 2014. Demand pattern calibration in water distribution networks. *Procedia Engineering*, 70, pp.1495–1504.

Sarrate, R., Blesa, J. & Nejjari, F., 2014. Sensor placement for leak detection and location in water distribution networks. *Water Science and Technology: Water Supply*, 14(5), pp.795–803.

Savic, D.A., Kapelan, Z.S. & Jonkergouw, P.M.R., 2009. Quo vadis water distribution model calibration? *Urban Water Journal*, 6(1), pp.3–22.

Savic, D.A. & Walters, G.A., 1995. *Genetic algorithm techniques for calibrating network models*, Exeter, UK. Available at: http://emps.exeter.ac.uk/media/universityofexeter/schoolofengineeringmathematicsandphysicalsciences/research/cws/downloads/Report_95-12.pdf.

De Schaetzen, W.B.F., 2000. *Optimal calibration and sampling design for hydraulic network models*. University of Exeter. Available at: http://emps.exeter.ac.uk/engineering/research/cws/research/distribution/design-hydraulic-network-.html.

Seaford, H., 1994. Acoustic leak detection through advanced signal processing technology. *Noise and Vibration Worldwide*, 25(5), pp.17–18.

Severn Trent Water Ltd, 2018. Thank you to all of our customers. *Severn Trent Water*. Available at: https://www.stwater.co.uk/news/news-releases/thank-you-to-all-of-our-customers/.

Sewerin, 2015. SeCorrPhon AC06. Available at: http://www.sewerin.co.uk/products/water-leak-location/secorrphon/ [Accessed November 26, 2018].

Shamir, U., 1974. Optimal Design and Operation of Water Distribution Systems. *Water Resources Research*, 10(1), pp.27–36.

Shamir, U. & Howard, C.D.D., 1977. Engineering Analysis of Water-Distribution Systems. *Journal of American Water Works AssociationAmerican Water Works Association*, 69(9), pp.510–514.

Shamir, U. & Howard, C.D.D., 1968. Water distribution system analysis. *Journal*

of the Hydraulic Division, ASCE, 94(1), pp.219–234.

Shinozuka, M., Liang, J.W. & Feng, M.Q., 2005. Use of supervisory control and data acquisition for damage location of water delivery systems. *Journal of Engineering Mechanics-Asce*, 131(3), pp.225–230.

Soldevila, A. et al., 2016. Leak localization in water distribution networks using a mixed model-based/data-driven approach. *Control Engineering Practice*, 55, pp.162–173. Available at: http://dx.doi.org/10.1016/j.conengprac.2016.07.006.

Soldevila, A. et al., 2017. Leak localization in water distribution networks using Bayesian classifiers. *Journal of Process Control*, 55, pp.1–9.

Song, M. & Chen, D.M., 2018. A comparison of three heuristic optimization algorithms for solving the multi-objective land allocation (MOLA) problem. *Annals of GIS*, 24(1), pp.19–31. Available at: https://doi.org/10.1080/19475683.2018.1424736.

Sophocleous, S. et al., 2016. A Graph-based Analytical Technique for the Improvement of Water Network Model Calibration. In *Procedia Engineering*. 12th International Conference on Hydroinformatics, pp. 27–35.

Sophocleous, S. et al., 2017. A model pre-processing approach for improving calibration-based leakage detection using a genetic algorithm Artificial Generation of Field Test Data. In Sheffield: Computing and Control for the Water Industry.

Sophocleous, S. et al., 2017. A Two-stage Calibration for Detection of Leakage Hotspots in a Real Water Distribution Network. In *Procedia Engineering*. 18th International Conference on Water Distribution Systems Analysis.

Sophocleous, S. et al., 2015. Advances in water mains network modelling for improved operations. In *Procedia Engineering*. 13th Computer Control for Water Industry Conference, pp. 593–602.

Sophocleous, S. et al., 2018. Leak Detection and Localization Based on Search Space Reduction and Hydraulic Modelling. In 1st International WDSA/CCWI 2018 Joint Conference.

Sousa, J. et al., 2015. Locating leaks in water distribution networks with simulated annealing and graph theory. *Procedia Engineering*, 119(1), pp.63–71. Available at: http://dx.doi.org/10.1016/j.proeng.2015.08.854.

South, M.C., Wetherill, G.B. & Tham, M.T., 1993. Hitch-hiker's guide to genetic

algorithms. *Journal of Applied Statistics*, 20(1), pp.153–175.

Speight, V.L., 2015. Innovation in the water industry: barriers and opportunities for US and UK utilities. *Wiley Interdisciplinary Reviews: Water*, 2(4), pp.301–313. Available at: http://doi.wiley.com/10.1002/wat2.1082.

Sreepathi, S. & Mahinthakumar, G.K., 2013. Optimus: A parallel optimization framework with topology aware PSO and applications. In *Proceedings - 2012 SC Companion: High Performance Computing, Networking Storage and Analysis, SCC 2012*. Salt Lake City, UT, USA: IEEE, pp. 1524–1526.

Storn, R. & Price, K., 1995. Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces. *Journal of Global Optimization*, 23(1), pp.1–15.

Suzuki, T., 1983. Uniqueness and Nonuniqueness in an Inverse Problem for the Parabolic Equation. *Journal of Differential Equations*, 47(2), pp.296–316.

Tabesh, M., Asadiyami Yekta, A.H. & Burrows, R., 2009. An integrated model to evaluate losses in water distribution systems. *Water Resources Management*, 23(3), pp.477–492.

Tabesh, M., Jamasb, M. & Moeini, R., 2011. Calibration of water distribution hydraulic models: A comparison between pressure dependent and demand driven analyses. *Urban Water Journal*, 8(2), pp.93–102.

Taghvaei, M., Beck, S.B.M. & Staszewski, W.J., 2006. Leak detection in pipelines using cepstrum analysis. *Measurement Science and Technology*, 17(2), pp.367–372.

Tanyimboh, T.T. & Seyoum, A.G., 2016. Multiobjective evolutionary optimization of water distribution systems: Exploiting diversity with infeasible solutions. *Journal of Environmental Management*, 183, pp.133–141. Available at: http://dx.doi.org/10.1016/j.jenvman.2016.08.048.

Tao, T. et al., 2014. Burst Detection Using an Artificial Immune Network in Water-Distribution Systems. *Journal of Water Resources Planning and Management*, 140(10). Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000405.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Philadelphia: Society for Industrial and Applied Mathematics.

The Guardian, 2017. Thames Water given maximum £8.5m fine for missing leak target. *Water*. Available at:

https://www.theguardian.com/environment/2017/jun/14/thames-water-given-maximum-fine-for-missing-leak-target [Accessed November 16, 2018].

Thornton, J., 2003. Managing Leakage by Managing Pressure: A Practical Approach. *Water 21*, pp.43–44. Available at: http://env1.kangwon.ac.kr/leakage/2009/knowledge/papers/pressure/Water Loss-Oct ABS.hwp.

Tikhonov, A.N., 1963. On the solution of ill-posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR*, 151(3), pp.501–504.

Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solutions of Ill-Posed Problems*, Washington DC: Winston and Sons.

Todini, E., 1999. Using a Kalman Filter Approach for Looped Water Distribution Network Calibration. In D. A. Savić & G. A. Walters, eds. *Proceedings of Water Industry Systems: Modelling and Optimisation Applications*. Exeter, UK, pp. 327–336.

Tucciarelli, T., Criminisi, A. & Termini, D., 1999. Leak Analysis in Pipeline Systems by Means of Optimal Valve Regulation. *Journal of Hydraulic Engineering*, 125(3), pp.277–285.

TWGWW, 1980. *Leakage Control Policy and Practice*, UK.

Unser, M., 2016. Characterization of the Solution of Linear Inverse Problems with Generalized TV Regularization. In *Imaging and Applied Optics*. Heidelberg, Germany: Optical Society of America.

USGS, 2011. The World's Water. *United States Geological Survey*. Available at: https://water.usgs.gov/edu/earthwherewater.html [Accessed November 16, 2018].

Vapnik, V. & Kotz, S., 1982. *Estimation of Dependences Based on Empirical Data: Empirical Inference Science (Information Science and Statistics)* 2nd ed., Springer-Verlag Berlin Heidelberg.

Vassiljev, A., Koor, M. & Koppel, T., 2015. Real-time demands and calibration of water distribution systems. *Advances in Engineering Software*, 89, pp.108–113. Available at: http://dx.doi.org/10.1016/j.advengsoft.2015.06.012.

Vela, A., Perez, R. & Espert, V., 1991. Incorporation of leakage in the mathematical model of a water distribution network. In *Proceedings of the*

2nd international conference on computing methods in water resources. Computational Mechanics Publication, pp. 245–257.

Vitkovsky, J.P. & Simpson, A.R., 1997. *Calibration and Leak Detection in Pipe Networks Using Inverse Transient Analysis and Genetic Algorithms*, Adelaide.

Vıtkovsky, J.P., Simpson, A.R. & Lambert, M.F., 2000. L d c u g a t. *Journal of Water Resources Planning and Management*, 126(4), pp.262–265.

Walski, T., 2000. Model calibration data: The good, the bad, and the useless. *Journal of American Water Works Association*, 92(1), pp.94–99.

Walski, T., Wu, Z. & Hartell, W., 2004. Performance of Automated Calibration for Water Distribution Systems. In *World Water and Environmental Resources Congress*. Salt Lake City, Utah, United States: American Society of Civil Engineers.

Walski, T.M. et al., 2008. Determining the Accuracy of Automated Calibration of Pipe Network Models. In *Eighth Annual Water Distribution Systems Analysis Symposium (WDSA)*. American Society of Civil Engineers.

Walski, T.M., 1995. Standards for model calibration. In *Proceedings of the 1995 AWWA Computer Conference*. Norfolk, pp. 55–64.

Walski, T.M., 1983. Technique for Calibrating Network Models. *Journal of Water Resources Planning and Management*, 109(4), pp.360–372. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%281983%29109%3A4%28360%29.

Walski, T.M. et al., 2001. *Water Distribution Modeling*, Haestad Press.

Walski, T.M., Sharp, W.W. & Shields, F.D., 1988. Predicting Internal Roughness in Water Mains. , 86(617), pp.0–4.

Wang, X. et al., 2001. Leak detection in pipeline systems and networks : a review. In *Conference on Hydraulics in Civil Engineering*. Hobart, Australia: The Institution of Engineers, pp. 1–10.

Water Authorities Association UK, 1980. *Leakage Control Policy and Practice*, London. Available at: http://whqlibdoc.who.int/hq/1997/WHO_EOS_97.03_(appendix1).pdf.

Water Briefing, 2013. Ofwat calls water cos PR14 quality assurance processes “not fit for purpose”. *Regulation and Legislation*. Available at: https://www.waterbriefing.org/home/regulation-and-legislation/item/8293-

ofwat-calls-water-cos-pr14-quality-assurance-processes-“not-fit-for-purpose”?font-size=smaller [Accessed November 27, 2018].

WRc, 1989. *Network Analysis - A code of practice*, Swindon, England: Water Research Center.

Wu, Z.Y. et al., 2002. Calibrating Water Distribution Model Via Genetic Algorithms. In *AWWA IMTech Conference*. Kansas City, Missouri, USA: American Water Works Association, pp. 1–10.

Wu, Z.Y., 2009. Unified parameter optimisation approach for leakage detection and extended-period simulation model calibration. *Urban Water Journal*, 6(1), pp.53–67.

Wu, Z.Y. et al., 2011. *Water loss reduction*, Exton, Pennsylvania, USA: Bentley Institute Press.

Wu, Z.Y. & Clark, C., 2009. Evolving effective hydraulic model for municipal water systems. *Water Resources Management*, 23(1), pp.117–136.

Wu, Z.Y. & Sage, P., 2006. Water Loss Detection via Genetic Algorithm Optimization-based Model Calibration. In *Annual International Symposium on Water Distribution Systems Analysis*. Cincinnati, Ohio, USA: American Society of Civil Engineers, pp. 1–11.

Wu, Z.Y., Sage, P. & Turtle, D., 2010. Pressure-Dependent Leak Detection Model and Its Application to a District Water System. *Journal of Water Resources Planning and Management*, 136(1), pp.116–128. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%290733-9496%282010%29136%3A1%28116%29.

Wu, Z.Y. & Walski, T.M., 2012. Effective Approach for Solving Battle of Water Calibration Network Problem. *Journal of Water Resources Planning and Management*, 138(5), pp.533–542. Available at: http://ascelibrary.org/doi/10.1061/%28ASCE%29WR.1943-5452.0000193.

Xie, X. et al., 2017. Compressed sensing based optimal sensor placement for leak localization in water distribution networks. *Journal of Hydroinformatics*, 20(6), pp.1286–1295. Available at: http://jh.iwaponline.com/lookup/doi/10.2166/hydro.2017.145.

Ye, G. & Fenner, R.A., 2011. Kalman Filtering of Hydraulic Measurements for Burst Detection in Water Distribution Systems. *Journal of Pipeline Systems Engineering and Practice*, 2(1), pp.14–22. Available at:

http://ascelibrary.org/doi/10.1061/%28ASCE%29PS.1949-1204.0000070.

Yeh, W.W., 1986. Review of Parameter Identification Procedures in Groundwater Hydrology: The Inverse Problem. *Water Resources Research*, 22(2), pp.95–108.

Yu, W., 1991. On the existence of an inverse problem. *Journal of Mathematical Analysis and Applications*, 157(1), pp.63–74.

Zabinsky, Z.B., 2011. Random Search Algorithms. *Wiley Encyclopedia of Operations Research and Management Science*.

Zhang, Q. et al., 2016. Leakage Zone Identification in Large-Scale Water Distribution Systems Using Multiclass Support Vector Machines. *Journal of Water Resources Planning and Management*, 142(11), p.04016042. Available at: http://ascelibrary.org/doi/10.1061/(ASCE)WR.1943-5452.0000661.

Zhou, Z.J. et al., 2011. Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection. *Expert Systems with Applications*, 38(4), pp.3937–3943. Available at: http://dx.doi.org/10.1016/j.eswa.2010.09.055.