INTEGRATED OPTIMAL PRESSURE SENSOR PLACEMENT AND LOCALISATION OF LEAK/BURST EVENTS USING INTERPOLATION AND A GENETIC ALGORITHM

SHAUN ROY BOATWRIGHT

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF ENGINEERING

THE UNIVERSITY OF SHEFFIELD
FACULTY OF ENGINEERING
DEPARTMENT OF CIVIL AND STRUCTURAL ENGINEERING

I, Shaun Roy Boatwright, confirm that the thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means. This work has not been previously been presented for an award at this, or any other, university.

…………………………………………………………………..

SUBMITTED DECEMBER 2020

Abstract

Leak/burst events are a serious problem because they disturb customer supplies, lead to water loss and managing them consumes vast resources. Water companies are continually seeking solutions to improve the situation. Presented in this thesis is the development, verification and validation of an integrated framework of methods for determining the optimal configurations of pressure sensors in a DMA and for localising new leak/burst events using a data-driven leak/burst localisation technique. The integration of the leak/burst localisation technique with the sensor placement technique is a novel feature of this framework of methods.

A data-driven leak/burst localisation technique, featuring a novel spatially constrained inverse-distance weighted interpolation technique, was developed which quantifies the change in pressure due to a new leak/burst event using pressure sensors deployed in a DMA, without using a hydraulic model. The leak/burst localisation technique combines data from multiple pressure sensors to localise a leak/burst event by interpolating using the distance travelled along pipes. The leak/burst localisation technique was combined with the GALAXY multi-objective evolutionary algorithm to identify the optimal sensor configurations and parameters for the leak/burst localisation technique efficiently. The sensor placement technique automatically determines the leak/burst event sizes for each DMA and groups them to minimise the number of leak/burst event scenarios which are considered.

The framework of methods was developed and verified iteratively using data from hydraulic models and a real DMA and validated using data from 20 engineered events conducted in two real DMAs in the UK. During validation, the sensor placement technique identified the optimal sensor configurations from a constrained subset of hydrants in each DMA. The agreement between the leak/burst localisation performance for the real and modelled engineered events demonstrated that the sensor placement technique can accurately predict the expected level of performance which will be achieved in a real DMA, particularly as the number of optimal sensors increases. Engineered events as small as 3.5% of the peak daily flow (6% of the average daily flow) were correctly localised with search areas containing as few as 12% of the pipes in a DMA.

Acknowledgements

First and foremost, I wish to extend my heartfelt and sincere appreciation to Professor Joby Boxall and Dr Steve Mounce, of the University of Sheffield, and Dr Michele Romano, of United Utilities, for their tireless patience, dedication, support and encouragement throughout all stages of my doctorate. The time and advice you supplied along the way were invaluable and I will be forever grateful for having you all as my supervisors. Special thanks must also go to Dr Michele Romano for facilitating both sets of engineered events and providing the data which formed a key part of the work in this thesis. My thanks also extend to Kevin Woodward for his input during the early stages of this project.

I wish to acknowledge the support of the innovation team at United Utilities for providing the funding that made this research possible. A special mention must also go to Laura Conroy and Sally-Anne Flanagan for their help during the time spent at United Utilities. I am grateful also to the Engineering and Physical Sciences Research Council (EPSRC) for the funding provided (research grant number EP/L015412/1).

I extend my thanks to the STREAM industrial doctoral centre for providing me with this opportunity. I wish to thank Dr Daniel Aiken, Dr Thomas Collin and Dominic Carver, my cohorts and friends, who listened to many quibbles and provided much needed entertainment and solidarity, particularly during our time in Cranfield and at the TSEL and challenge weeks every year.

I am indebted to the counselling service at the University of Sheffield for the support provided during the final two years of my EngD. In particular I wish to recognise the care provided by Paul Hamnett who shone a light during the darkest times and without whom I could not have continued towards this goal.

Finally, I wish to acknowledge the support and love of my Mum, Anne, who never stopped believing in me and whose incessant support throughout my life has been vitally important. Likewise, to my brothers Karl and Glen, who kept my spirits up and provided endless hours of advice and humour when it was needed most.

Publications

The following conference paper contains work from this thesis:

Boatwright, S., Romano, M., Mounce, S. R., Woodward, K. and Boxall, J. B. (2016). Approximate location of leaks and bursts in a district metered area using statistical process control and geostatistical techniques, Proc. Computing and Control for the Water Industry Conference, Amsterdam, Netherlands.

This conference paper contains results covered in Chapter 5.

List of Figures

List of Tables

List of Abbreviations

AI	Artificial intelligence
AMP	Asset management plan
ANN	Artificial neural network
ART	Advanced reflection techniques
AWWA	American water works association
BIS	Bayesian inference system
CCW	Consumer council for water
CMP	Critical monitoring point
CFPD	Comparison of flow pattern distributions
CUSUM	Cumulative sum
CV	Check valve
DM	District meter
DMA	District metered area
DMZ	Demand monitoring zone
DPA	Discrete pressure area
DWI	Drinking water inspectorate
DWT	Discrete wavelet transform
EA	Environment agency
ELL	Economic level of leakage
EPR	Evolutionary polynomial regression
EWMA	Exponentially weighted moving average
FAVAD	Fixed and variable area discharge
FCV	Flow control valve
FIM	Fisher information matrix
FRF	Frequency response function
GA	Genetic algorithm
GIS	Geographical information system
GPR	Ground penetrating radar
IDW	Inverse-distance weighted
IRF	Impulse response function

ITA	Inverse transient analysis
IWA	International water association
LM	Levenberg-Marquadt
LNC	Leak noise correlator
LP	Local polynomial
MNF	Minimum night flow
MOEA	Multi-objective evolutionary algorithm
MSE	Mean squared error
NPW	Negative pressure wave
NRW	Non-revenue water
OC	Ordinary Cokriging
ODI	Outcome delivery incentive
OK	Ordinary Kriging
PCA	Principal component analysis
PIG	Pipeline inspection gauge
PRV	Pressure reducing valve
PSO	Particle swarm optimisation
SC-IDW	Spatially constrained inverse distance weighted
SIM	Service incentive mechanism
SOM	Self-organising map
SPC	Statistical process control
SVM	Support vector machine
TCV	Throttle control valve
TDNN	Time delay neural network
TDR	Time domain reflectometry
UFW	Unaccounted for water
WDS	Water distribution system
WSZ	Water supply zone

Nomenclature

W_{ij}	Weight between a measured and unmeasured location
d_{ij}	Distance between a measured and unmeasured location
n	Interpolation exponent
z_j	Estimated value at a single point, j
z_i	Measured value at a single point, i
x_{sa}	Search area threshold absolute value
$threshold_{sa}$	Search area threshold
$Z_{j,min}$	Minimum interpolated value
$Z_{j,max}$	Maximum interpolated value
B_n	Bell number
k	Number of groups of leak/burst events
n_{lb}	Number of leak/burst event locations
C	Number of potential sensor configurations
s_p	Number of possible sensor locations
s_d	Number of sensors to be deployed
n_{adv}	Number of additional decision variables
n_{vdv}	Number of values for each decision variable
C_{dv}	Number of combinations of parameters

Table of Contents

Chapter 1 – Introduction

The supply of clean drinking water has been a primary concern of human civilisation for the last several thousand years. Providing a fixed means of supplying population centres with water meant that vital activities such as agriculture and sanitation were possible and allowed people to live more comfortably in close proximity. The rapid growth of the population of the Earth within the second half of the 20th century and early 21st century has led to increased scarcity of supply in many regions of the Earth. This coupled with the predicted rise in the Earth's temperature and corresponding rise in sea levels means that even less freshwater will be available due to saltwater intrusion into both ground and surface water sources. Water in most developed countries is collected from aquifers, rivers and lakes, treated at central facilities and distributed by means of fixed conduits, such as pipes, to the population of a given geographic area. The size and capacity of the water distribution system (WDS) is dictated by the demographics of the supplied area and the distance between the source and the served population. Water is inevitably lost between the points of treatment and consumption due to failures in the WDS. Firstly, WDSs are generally large, complex structures primary located below other vital infrastructure. This means access for inspection and maintenance is very disruptive and expensive which discourages water companies from performing regular maintenance. Secondly the buried infrastructure can be up to 100 years old with no historical records of its condition or even exact location available. This further increases the challenge of planning preventative maintenance and increases the likelihood of failure. Lastly, water companies do not want to spend money performing preventative maintenance as it is very hard to predict which sections of WDS are most likely to fail and hence should be replaced immediately.

Failures of the pipes in WDSs can have serious consequences for water companies as they lead to loss of water (or leakage). Estimates vary between regions of the world but water loss in developed countries averages around 15% of production whilst this figure is roughly 35% for developing countries (Kingdom et al., 2006). Treating water for consumption is expensive in terms of financial cost, energy, carbon footprint and chemical demand and water companies must spend money treating the water that is lost from the WDS without recovering the cost from customers. Also of concern are regulatory drivers which aim to limit leakage by placing financial penalties on water companies for poor performance against targets (Alegre et al., 2013). The UK water sector is highly regulated in terms of both quality of supply (Drinking Water Inspectorate, DWI), economic competition (Ofwat) and the environment (Environment Agency, EA). Since privatisation in 1989, the UK water companies have been set economic and quality performance targets (asset management plans (AMPs)) every five years to ensure that competition is maintained between the companies. This is because customers are supplied by location and do not have the opportunity to switch supplier as they would for other utilities. Part of the review process involves agreeing to a set of outcome delivery incentives (ODIs) which measure progress against a set of pre-defined objectives with rewards and penalties in place for over and underperformance, respectively. Water loss from WDSs directly affects the following ODIs (i) Water Quality Supply Index, (ii) Average Minutes Supply Lost per property per year, (iii) Reliable Water service index, (iv) Security of supply index and (v) Total leakage at or below target. The total financial impact that these five ODIs can have over the course of one review period is significant to the water companies and shows how important managing this is.

Far from being purely an economic, regulatory issue pipe failures and the resultant low pressures in the WDS can have far reaching consequences in terms of the quality of water supplied. Failures of water pipes can also introduce contamination into the network, leading to customer complaints. Usually this would be on the grounds of colour, taste or odour which are the most noticeable to consumers but in extreme cases this could lead to the introduction of harmful contaminants such as chemicals, pesticides, bacteria and other harmful organisms (LeChevallier et al., 2003). There are also number of environmental impacts associated with treating and transporting the water lost from the network including increased carbon emissions and pressure on water sources. Water companies frequently have to find alternative sources for the water that is supplied to customers as abstraction licenses are revoked or reduced in an effort to conserve natural ground and surface water sources by the EA. This is especially prevalent in arid regions with high temperatures where water would typically have to be moved great distances from areas with surpluses (NRDC, 2012). Reduced water loss can lead to the postponement of the development of new sources, placing far less strain on an already pressured resource. On top of this, water consumers have strong opinions regarding leakage even if it does not affect their level of service or the quality of the water that they are supplied. Based upon research conducted by the Consumer Council for Water (CCW), 22% of customers thought that reducing leaks was the biggest challenge for the water industry in the UK and Wales over the next ten years (SPA Future Thinking, 2013). Leakage levels affect how customers perceive their water company and impacts how likely they are to engage in water saving schemes, which can have a big influence on water demand within a WDS.

It is worth noting that all water networks experience loss in one form or another but the levels and types of loss will vary significant around the world and even in certain countries. Whilst reducing water loss provides many benefits it is also highly expensive leading to a trade-off between the benefits and costs of doing so (Pearson and Trow, 2005). The economic level of leakage (ELL) defines the percentage water loss for a given water network which leads to the minimum cost when summing the cost of fixing and leaving leaks. This means that water companies should be aiming to maintain leakage at the ELL. It is worth pointing out that the ELL is not a static point but develops over time as the costs of leakage and reducing leakage change. The cost of reducing water loss is falling as more advanced strategies and technologies are conceived for detecting and repairing the events which lead to water loss in WDSs but this is often counteracted by the increased costs of water loss in terms of the aspects already describes so far. In recent years the amount of monitoring of WDS has increased as it is understood that collecting more information allows better strategic planning of maintenance interventions. Of particular interest to water companies is real or near-real time collection and analysis of data to allow them to respond more quickly to operational problems within the network. This can mitigate the impact these problems have on the ODIs, thereby reducing the financial risk to a water company. Water companies do not currently make full use of the data that is collected because of the high costs associated with processing and analysing that data and the uncertainty regarding the benefits that could be seen. As regulatory budgets tighten water companies need to develop new methods to analyse the ever-increasing amounts of data to draw meaningful conclusions about where money should be invested within a WDS to promote efficiency.

Chapter 2 - Literature Review

2.1. Introduction

It is important to define unambiguously what is meant by water loss and where it can arise within a WDS. The international water association (IWA) (IWA, 2000) and American Water Works Association (AWWA) have a well-defined terminology (AWWA, 2003) which is used throughout the international water sector to define the components of water loss. All types of non-revenue water (NRW), also called unaccounted for water (UFW), represent a financial loss and a threat to the security of supply to water companies and are therefore undesirable. The largest of these in terms of volume are real losses which is the physical phenomenon of losing water from the network between treatment and consumption (including leakage from transmission and distribution mains, leakage and overflow at storage tanks and leakage on service connections between the transmission pipe and the customer meter). Apparent losses also account for a proportion of water losses and include metering inaccuracies and unauthorised consumption from illegal connections or hydrant use. There are different methods by which water loss can be managed within a WDS and these will be discussed further in the remainder of this chapter.

It has always been known that a WDS can lose water but historically there was not much emphasis on trying to resolve the issue. This was due to the lower demand pressures on water resources, lack of environmental regulation and the fact that the average age of water assets was lower. In recent years there has been greater appreciation of the mechanisms which cause water loss from WDS and how this can be prevented. Early studies in the field were primarily focussed on pipeline integrity and not the further impacts this has. Andreou et al. (1987) undertook a case study analysing WDS pipe break records for six utilities in the United States in order to understand the mechanisms of failure. This was then linked to economic interventions which could be made depending upon the type and location of the failure. According to the IWA, real losses can also be classified by their magnitude into (i) background leakage (ii) reported leaks/bursts and (iii) unreported leaks and bursts (IWA, 2000).

Generally, individual background leaks will have a small flow rate and little impact on the WDS and service levels experienced by customers. Typically, they will have a long run time as they are difficult to detect so the resulting water loss from background leaks can be high. Reported leaks will have higher flow rates which will cause impact to customers and hence will be fixed quickly but if sufficiently large they can lead to extended loss of supply and negative publicity for a water company. Lastly the unreported bursts will not impact upon the service levels but are large enough to detect and rectify if a suitably sensitive leakage control policy is in place.

The size and complexity of a WDS means that monitoring their integrity to try and predict the location of failures prior to them happening is nearly impossible. This is further exacerbated by the budgetary constraints which exist for most water companies. There exists a balance between spending money on reducing water loss and the value of the water which is lost. To minimise the total cost of water loss it is not optimal to reduce it zero as this would be prohibitively expensive. Instead, there is a point of minimum cost, the ELL, which balances the cost of water loss against the cost of reducing it. Maintaining the ELL ensures that the cost of water loss is minimised which is the most responsible way to manage the problem. The ELL is impacted by regulations, the available technologies to combat water loss and the availability of water, amongst other things. New schemes for water loss reduction must be economic in terms of allowing a water company to move towards the ELL (Pearson and Trow, 2005).

The most promising techniques have been based upon detecting, locating and repairing events once they have occurred to try and minimise the impact on customers and the associated water loss. Historically, water companies have relied on customers reporting deficient levels of service or problems with their water supply which ensures that resources are only deployed when they are definitely needed. This approach is efficient but is also problematic given the highly regulated nature of the industry. Customer service plays an ever-increasing role in evaluating water company performance in the UK through the use of, for example, the service incentive mechanism (SIM). Reducing the number of customer contacts by proactively managing and reporting service problems to customers prior to them contacting their water company will bring financial, regulatory and reputational benefits. Reactive maintenance is also expensive and disruptive as repairs are often charged using emergency tariffs and road closures are required in order to access the pipes in a WDS to perform the necessary repairs. Background leakage is particularly problematic as it will not lead to customer contacts meaning its presence cannot be determined easily.

There are several classes of techniques which are used to aid water companies in their aim to reduce non-revenue water (NRW). As discussed by Puust et al. (2010) leakage management can be broken down into (i) assessment, (ii) detection and (iii) control. Assessment is used to provide a picture of the performance of the network for regulatory and planning purposes but cannot provide information relating to specific portions of the network or to specific times at which problems might be occurring. Detection aims to confirm the presence of network events and, where district metered areas (DMAs) have been implemented, to provide an approximate location so that targeted intervention can be undertaken to resolve the problem quickly and with minimal disruption to customers. Leakage control covers the activities in which leaks are discovered and fixed and is the mechanism by which the water loss is physically reduced. These methods are used in combination to provide the most economical solution to the problem of NRW management. There has necessarily been a shift in attitudes within the industry to reflect the changing regulatory environment. In order to maximise the efficiency of running their infrastructure, asset management is now the favoured approach for water companies. This usually requires far greater levels of information to be available (and in greater detail/resolution) about assets than previous approaches so that decisions made are properly informed with high levels of confidence. The amount of instrumentation in WDSs is greater than ever and higher volumes of data are being collected as a result. This data is not always analysed however due to the labour-intensive nature, time and cost of doing so. Valuable insights into the operation of WDSs are often missed because of this which leads to the potential of automated, fast and reliable techniques which can transform this data into information which can be used to drive strategy within water companies.

In this section a comprehensive review of the available literature will be undertaken to establish the current state of knowledge and practices within the water industry. A number of techniques are currently used to allow the assessment, detection and control of leaks within WDS. Previous reviews by Puust et al. (2010), Mutikanga et al. (2013), Colombo et al. (2009) and Farley and Trow (2003) cover the range of various techniques and how they can be classified.

Water companies spend a great amount of time and resources trying to evaluate leakage within their WDS. This is a regulatory requirement and must be reported to Ofwat (for UK water companies) every year to benchmark water companies against their targets and each other. This also serves as an aid to plan the amount of expenditure which will be required in order to maintain performance levels. Assessment allows targeted interventions to be made by water companies so that any resources can be allocated where they will have the most benefit but without having knowledge of the presence of any individual water losses. Within the UK, and most developed water supply infrastructures, the practise of dividing the network into DMAs allows a localised water audit based upon the minimum night flow (MNF). DMAs were developed and implemented in the UK in the early 1980s (Morrison et al. 2007). A DMA is a sub-section of the network which is partially hydraulically isolated from the surrounding network by use of boundary valves and flow is monitored at all entry and exit points. Any increases in the MNF over time would be registered on the flow meter and could indicate that there are leak/burst events (or other events) in that DMA which allows a water company to deploy operatives to try to find and repair the event. A development of this is pressure measurement within DMAs to ensure that the minimum pressure at the customers stop tap of 7m (Ofwat, 2017) is maintained to all properties. This is usually measured at the point of highest elevation in the network or the point furthest away from the inlet to the DMA, either of which can be termed the critical monitoring point (CMP).

The range of available techniques for managing leakage can be divided into hardware and hydraulic techniques. Currently hardware-based techniques dominate the range of practical methods which are used by the utilities in their leakage management activities. This is because they have high accuracy, precision and reliability. However, they are highly labour intensive and hence expensive. These techniques include leak noise correlators (Hunaidi et al. 2000) acoustic sensors mounted on pipeline inspection gauges (PIGs – Grigg, 2006), and ground penetrating radar (GPR – Metje et al. 2007). The primary use of the hardware-based technique is to allow the discovery and repair of individual leak/burst events, in turn reducing the amount of leakage seen from a given DMA. This most commonly involves “sounding” pipes using a leak-noise correlator to discover the sources of leak/burst event noise and calculate the location.

Hydraulic techniques are also used by water companies. The main techniques used are water audits (AWWA, 2003), step testing (Boulos and Aboujaoude, 2011), steady state analysis-based techniques (Puust et al. 2010), transient analysis-based techniques (Colombo et al. 2009), negative pressure wave-based techniques (Ge et al. 2008) and data-driven statistical/artificial intelligence-based techniques (Wu and Liu, 2017). Of these, water audits and step testing are standard practise in the UK, as assessment and detection techniques due to their simplicity and effectiveness. However, in recent times more effort has been devoted to developing near real-time hydraulic methods for analysing pressure and flow data collected from a WDS in order to allow timely interventions to be made.

In the last decade artificial intelligence (AI) and statistical techniques have been developed to allow rapid detection of leak/burst events occurring within a WDS. A number of these approaches attempt to utilise the increased amount of data to gain further insight into the nature of a leak/burst event (such as magnitude or location) to allow operational interventions to be made in the most effective way possible. Numerous techniques based upon modelled leak/burst events have been proposed but techniques utilising data collected from a real WDS are less advanced because they require instrumentation to be installed and obtaining a usable set of leak/burst event data is more difficult and invasive. Studies such as Mounce et al. (2003), Romano et al. (2014a, 2014b), Srirangarajan et al. (2010a, 2010b) and Misiunas et al. (2005) demonstrate how the collected hydraulic data taken from the network can be analysed to detect and localise leak/burst events within a DMA. In previous reviews by Puust et al. (2010), Mutikanga et al. (2013) and Li et al. (2015) it was noted that the most effective strategy would likely be found by using a hydraulic technique to reduce the possible search area prior to using a hardware-based technique which would subsequently be used to locate a leak for repair. Also highlighted was the lack of research which has been undertaken using real leak/burst event data and a lack of applicability of the current research methods because they are mainly developed/tested using hydraulic model and not real leak/burst event data.

2.2. Hardware Based Techniques

2.2.1. Overview

Hardware-based techniques are a broad collection of methods which can allow the location of leak/burst events within a WDS to be ascertained accurately enough so that a repair can be carried out. They generally measure one or more hydraulic variable from the pipes in a WDS and the location of leak/burst events can be inferred by detecting a significant change in one or more measured variable which occurs as a result of the presence of a leak/burst event.

The techniques which have been found in the reviewed literature are the following which are subdivided as outlined by Romano et al. (2012):

- Acoustic techniques
- Tracer gas
- Thermographic techniques
- Ground penetrating radar techniques
- Inline pipe inspection gauge (PIG)

Here a review of each group of techniques is provided along with a discussion of the applicability and limitations of each of the techniques.

2.2.2. Acoustic techniques

Acoustic techniques allow detection and location of leak/burst events in a WDS by measuring the noise that is generated by water leaving the pipe. The amplitude and frequency of acoustic signals which are generated by leak/burst events is dependent upon the fluid in the pipe, the pressure within the pipe, the pipe size and material, the condition of the ground in which the pipe is buried and the size/type of leak/burst event (Wang et al. 2001). The sound wave which is generated by a leak/burst event will propagate through the fluid, pipe wall and ground surrounding the pipe and can be remotely detected by measuring the generated sound waves either on the pipe or in the ground surrounding the pipes. Acoustic sensors generally can be subdivided into three categories; leak noise correlators, geophone/hydrophone and noise loggers. Though each of the categories of sensors are designed to measure similar changes in the noise in a WDS there are variations when using each of the sensor types which will be explained in this section.

2.2.2.1. Listening Devices

Early acoustic methods for locating leaks were “listening sticks” which were extended wooden rods which allowed the vibration in the network to be heard by an operative. Electronic devices are now available which are much more sensitive than the previously used manual methods (Pilcher, 2003). Geophones and hydrophones can be used to listen to (or record) sound or vibration which is propagated through pipes and the contained water. Geophones are manual devices which are placed on pipe sections or fittings to measure the level of vibration. The presence of a leak/burst event will be accompanied by an increase in vibration in the pipe adjacent to the leak/burst event allowing its location to be inferred. Similarly, hydrophones measure the noise levels within the fluid and this is used to infer the presence and location of a leak/burst event. They are more sensitive than geophones and can detect leak noise up to 600m away whereas geophones are limited to roughly 250m (De Silva et al., 2011). The main drawback to the use of these listening devices is that they are very dependent on having an experienced operator and even then the interpretation of leak/burst noise from traffic and background noise can be very difficult. Additionally, the concentration and noise perception of the operator will decrease over time and leak/burst events can be missed.

2.2.2.2. Leak Noise Correlators

Leak noise correlators (LNCs) were first developed in the late 1970s (Grunwell and Ratcliffe, 1981) although they were of little operational use to water companies until the late 1980s. Much development has occurred since then which makes them a staple in the leak/burst event detection and location activities of water companies around the world. Many commercial LNCs have been developed such as LOKAL (Fuchs and Riehle, 1991). Early problems such as user subjectivity and difficulty when being used to detect and locate leak/burst events in large diameter plastic pipes, which represents the worst-case scenario, have been partially overcome in the last 15 years.

LNCs rely on the analysis of sound signals measured at multiple points within a pipe network (usually in fairly close proximity). The arrival time of the acoustic signals from a leak to each of the sensors depends on the propagation velocity of acoustic waves in water filled pipes. Multiple parameters such as the diameter of the pipe, the pipe wall thickness, fluid bulk modulus and the young's modulus of the pipe affect the time of arrival of a sound wave from a leak/burst event to the LNC. Errors in the propagation velocity of the waves will lead to localisation errors so the velocity must be confirmed to ensure accuracy. In order to improve the accuracy of the leak/burst location, built-in algorithms have been developed for reduction of interference noise which greatly improves the effectiveness of this method. Examples of these techniques being utilised on a real WDS can be found in (Hunaidi et al. 2004).

One study utilising the LOKAL method on a real WDS showed that the cost of the amount of water saved outweighed the cost of implementing the leak detection scheme. In practice, surveying large sections of a WDS is extremely time consuming and using LNCs should be targeted by using a combination of online monitoring and detection/location to minimise the cost incurred when using LNCs by a reduction of the search area. LNCs are simpler for operatives to use and require less complex interpretation of the acoustic signals which can generally be automated by the device. The drawbacks of using LNCs are that in order to obtain a precise leak/burst event location system knowledge is required to correlate the propagation velocities correctly, and for small leak/burst events (with low flow rates) the location can be difficult to calculate. In order to locate small leak/burst events the sensors must be placed very close to the leak further increasing the time taken to survey a given area of the WDS.

2.2.2.3. Noise Loggers

Noise loggers are permanent devices, placed approximately 200-500m apart (Hunaidi and Wang, 2006), which are installed on the pipes in a WDS to allow the noise to be monitored and analysed in a continuous fashion. Noise can be obtained continuously but the most efficient way to practically measure noise is during the period of lowest flow where the leak noise is highest when compared to the background noise. Historically they have been widely used within gas networks as they are effective on networks with simple topologies and they are economically viable against the high cost of pipe failures in gas networks (Loth et al. 2003). Recently they have become more prevalent within the water industry as the cost of instrumentation and data storage continues to decrease for water companies and higher financial penalties are placed on leakage targets. However, Hunaidi et al. (2004) highlighted that permanently deployed noise loggers had a payback period of 25 years, which is very high for water company assets and also that temporarily deployed loggers failed to detect almost 40% of leaks found in detailed listening surveys. Loggers can be used where a detailed listening survey cannot be performed due to high ambient noise. It is clear that the use of acoustic techniques is effective and efficient for pinpointing leaks once they have been approximately located but that they are not viable as a stand-alone solution. This is primarily due to the high capital cost and labour requirement when performing searches over large areas. As leakage penalties increase and the economic level of leakage is driven lower by improved technologies then permanently deployed acoustic monitors will become viable.

2.2.3. Tracer Gas-Based Techniques

Tracer gas-based leak detection techniques have been used to detect leaks in WDSs but are not common. It involves the introduction of a non-flammable gas, such as hydrogen or helium, mixed with air to the WDS pipes which have been depressurised and isolated. Any gas which escapes from the pipe is detected at ground level by the use of gas detectors and leak/burst events can be located by analysing the distribution of concentrations in the area around the section of pipe being tested. The method is useful as it does not require direct access to the underground pipes (except for the point of gas injection), it can find multiple leak/burst events in a given pipeline section and it works equally well on all pipe sizes/materials. Hunaidi et al. (2000) investigated the possibility of detecting leaks whilst the pipes were still full of water to avoid the problems of supply interruptions and increase the speed of the technique. Problems were noted with the detection radius which was only 1m. This means that applying this technique to large sections of the WDS would be time consuming as the operator would still have to walk all of the pipes and leak/burst events could easily be missed. This method is also dependent upon weather conditions as the gas measurement will change depending upon wind direction.

2.2.4. Thermographic Techniques

Thermographic techniques utilise infrared radiation to detect anomalies in the heat distribution in the ground surrounding the pipes in WDSs. Water escaping into the ground from a leak/burst event will cause a localised reduction in the temperature indicating that a leak/burst event is present. Thermography is widely used in residential situations to pinpoint leakage under floorboards and within walls which are not readily apparent. This has been applied to detecting and pinpointing leaks in underground water mains. In Montreal, Canada (Fahmy and Moselhi, 2009) the method gave good accuracy when studying infrared images from a real WDS. In Doha, Qatar (Atef et al., 2016) ground penetrating radar (GPR) was used to identify the pipes below the ground and infrared images were overlaid and analysed to predict the location of the leaks. However, the time taken to collect and analyse the images would not be competitive with using acoustic methods and the range of this method is very much smaller than acoustic techniques which would lead to increased resource costs. This method is not applicable in adverse weather conditions and requires specialist cameras and software to capture and analyse the required data to allow the location of leaks.

2.2.5. Ground Penetrating Radar Techniques

GPR has been used for the detection of underground utility lines in the construction industry. This allows excavations to take place with a much lower risk of colliding with underground assets such as water pipes, electricity cables, sewers, gas pipes or communications cables (Caldecott et al., 1988). Many commercial GPR systems are available, although the development of GPR methods for leak detection is relatively recent in comparison. Studies by Demirci et al. (2012), Lai et al. (2016) and Stampolidis et al. (2003) employed GPR to detect and locate leak/burst events in experimental studies, which were based upon laboratory work. According to the literature there is a limited amount of application to real WDSs for locating leak/burst events using GPR. This is mainly because of the high capital cost of the equipment required which limits the applicability of the technique to water companies who typically have limited budgets for leak/burst event detection and location.

2.2.6. Inline Pipeline Inspection Gauges

Pipe inspection gauges (PIGs) are a well-established technique and have been used for cleaning and inspecting pipelines in the water, oil and gas industries (Fletcher and Chandrasekaran, 2008). They are usually made from a flexible material which is sized very similarly to the internal diameter of the pipe being cleaned to allow for the propulsion of the PIG by the conveyed fluid. More recently "smart" PIGs have been utilised to collect information on the internal condition of the pipe and assess whether there are defects or failures present in the pipe wall. The PIGs utilised in a study by Licciardi (1998) used a tethered PIG which was drawn incrementally through a 200mm pipe. The method was effective at detecting the induced leak/burst events although errors of up to 62% were found when calculating the flow rate.

PIGs are effective at localising leak/burst events and can, in some cases, pinpoint its orientation in the circumference of the pipe. They are however invasive and disruptive, potentially leading to shutdowns and contamination. PIGs require a point of entry and removal from the pipes meaning that not all sections of a network can be surveyed in this manner. There is a limitation on the minimum pipe size which can pass a PIG and pipe fittings/valves may act as obstructions. Contact between the PIG and the pipe will dislodge material which could potentially end up at the customer taps causing issues with turbidity or taste. Also, the slow speed of using PIGs compared to other available methods indicates that this method is unlikely to be financially viable for the vast majority of water companies in all but a small number of specialised situations.

2.3. Hydraulic Techniques

As opposed to the hardware-based methods, hydraulic variables (such as pressure and flow) can be measured and analysed to allow detection and localisation of leak/burst events within WDSs. It is becoming increasingly common for water companies to monitor their WDSs in order to gain insight into their operation to improve their performance. Hydraulic techniques (which analyse hydraulic data collected from WDSs) range in complexity from simple water audits to statistical/AI techniques which can, in near real-time, detect and localise leak/burst events in WDSs.

This section will provide a comprehensive review of the methods which have been developed to analyse the data which is being collected from a WDS. This section will give an overview of the operating principle of each method along with a summary of the advantages and disadvantages of each. The types of methods which have been reviewed are divided as follows as per Romano et al. (2012):

- Water audits
- Step testing
- Steady-state analysis-based techniques
- Transient analysis-based techniques
- Negative pressure wave-based techniques
- Statistical/AI-based techniques

2.3.1. Water Audits

A water audit (also commonly called a water balance) aims to identify all components of the water entering and leaving a WDS (or portion thereof) to provide an accurate estimate for how much water is lost. This methodology can utilise measurements and/or estimates of certain water volumes within the water balance depending on the level of accuracy required and the amount of resources which are available. The IWA (IWA, 2000) and AWWA (AWWA, 2003) have developed guidelines for performing water audits which includes a breakdown of the system input into various categories, depending on where and how they are lost.

The water balance can either be undertaken at the WDS scale or for individual DMAs. It is preferable, if possible, to audit each DMA individually in order to assess where within the network NRW is more of a problem. This will allow the worst performing areas of the network to be targeted and interventions taken if this is deemed appropriate. The water balance can be conducted from the top-down or bottom-up depending upon the level of available information which a water company has at its disposal. The top-down approach aims to quantify NRW (and in particular real losses) using estimations or high-level measurements for the input volume and authorised consumption. Using this and having knowledge of the apparent losses, such as meter inaccuracies and unauthorised consumptions, the real losses can be calculated. Generally, the top-down approach is quicker, less labour intensive and requires the collection of less information to be performed and can provide a good estimation of real losses to a water company.

Once the real losses have been calculated the water company can try to attribute them to different sections of the network (i.e. transmission mains, distribution mains or storage facilities) and assess which component is the priority for further assessment or reduction. It is worth highlighting that this assessment methodology will produce an estimate which a water company can use to make operational decisions, to assess the efficiency of their water supply operations and to plan strategic actions in order to improve this. Following the top-down assessment, which can be repeated until sufficient accuracy is achieved, bottom-up assessments aim to identify the individual components of real losses in a WDS (i.e. individual leak/burst events) in order to find the total. Individual components can be calculated in a number of ways but usually result from an understanding of the WDS operation and will be summed to try and bring them in line with the order of magnitude of the top-down approach. Measurements, usually flow and pressure, can be taken from a WDS (or different sections thereof) and analysed on a daily basis to find discrepancies which can be correlated to the presence of leak/burst events or changes in operation. This aims to assess the amount of leakage within a network (or section of the WDS) by analysing the flow data between 02:00 and 04:00 when the legitimate demands are at their lowest and leakage makes up a greater proportion of the demand. Leakage is a pressure dependent phenomenon according to the fixed and variable area discharge (FAVAD) principle and the increased pressure during the MNF causes the leakage flow to increase making it easier to detect when analysing the flow into the network or a DMA. In theory, a refined assessment will lead to similar numbers for both the top-down and bottom-up approaches although this can be difficult to achieve. Multiple passes of both assessments may be required until they agree well with each other. It is worth stating that for the bottom-up approach to be possible the network must be composed of DMAs which have flow measurement at the inlet and well defined, hydraulically isolated boundaries. Whilst this is the standard within the UK and for many other European companies it is by no means universal.

There are problems with using both the bottom-up and top-down methodologies. This is primarily because of the number of variables that need to be calculated in order to achieve high levels of accuracy and the difficulty of getting information required to calculate all of the components of the water balance. Difficulties arise from the collection and handling of vast quantities of data which leads to errors when performing the water balance. One source of this error is the meters which are registering the flow in to (and out of, where applicable) the DMA. Meter errors can be as high as 30% (Phair, 1997) which can lead to over or under registration of the flow into the DMA and hence will lead to inaccuracies in the predicted losses seen within a DMA. Another obstacle to calculating the NRW is the difficulty in estimating the unauthorised consumption. By their very nature water companies will not be aware of their presence or location and estimating the quantity of water lost is very difficult because of this (Liemberger and Farley, 2004). Water audits are useful to water companies because they are simple and can be adapted to suit the requirements of the users with ease. A number of case studies were reviewed by Klingel and Knobloch (2015) which shows the variety of approaches which have been taken by various water companies in different regions around the world. They are broken down into both developing countries and more developed countries and the trend is generally that the more developed countries tend to calculate more components of their water balance, presumably as there is more information available to them to do so. Having said this, it still shows that a lot of more developed countries are not determining some of the components of their water balance or are having to estimate them meaning that there are still gaps in the knowledge they have about their WDS which has the potential to limit the efficiency gains that can be made.

2.3.2. Step Tests

Step testing involves the systematic isolation of sub-sections within a DMA using valves (Pilcher et al., 2007). As the valves within the DMA are being progressively closed to artificially reduce the size of the DMA (i.e. starting with the valves furthest away from the inlet flow meter and working towards the meter) the flow on the inlet meter is monitored to see how it changes. Any large reduction in the flow on the meter, which could not be attributed to legitimate usage within a given isolated section of the DMA, indicates that there is a high level of leakage which should be investigated. Generally, this procedure will be carried out between 02:00 and 04:00 to coincide with the period of MNF. This will ensure that leakage makes up its maximum proportion of the flow into the DMA. It is recommended that no fewer than 10 steps are used as this will not provide sufficient localisation of the high leakage areas required to allow repairs or further investigations to take place.

A variation of this technique is to divide the DMA into several sub-areas with sub-meters installed at the inlet of each. The DMA will be divided into areas of approximately 300-500 properties each of which will log the flows into each of the sub-areas. This means that areas of high leakage can be identified more precisely than when considering an entire DMA and remedial action can be taken to address leakage in the worst performing sub-areas of a DMA. For both methods a reliable and accurate estimate of the night-time demand is required to be able to determine, indirectly, the leakage flow. Due to the stochastic nature of the water demands (also called consumptions) this is not easy and there will be some error when calculating how much leakage there is within a given isolated section of the network however this method can provide a good approximation for leakage managers who want to know where to target their resources most effectively within a WDS. These methods are not ideal as they are labour intensive and there is the risk of leaving parts of the network without supply for hours. The tests are performed during the MNF to minimise the risk of interrupting supply to customers which can have an impact on the regulatory performance of a water company.

2.3.3. Steady-State Analysis-Based Techniques

Steady-state analysis techniques have been developed which make use of measured network characteristics (such as pressure or flow) to allow verification of some unknown characteristics (e.g. pipe roughnesses). Steady-state conditions are achieved hydraulically once conditions of a fluid do not deviate significantly with time. In reality all fluids are inherently unsteady but this is only obvious when using high-precision instruments. Network characteristics, such as pipe roughnesses or demands at nodes, within a hydraulic model of a WDS can be tuned to produce the desired outputs (i.e. minimising the difference between measured and output values from a hydraulic model) which will usually be pressure or flow in order to understand how the network is operating.

Pudar and Liggett (1992) introduced the concept of steady-state analysis and a number of studies have been undertaken since to try to improve the procedure and increase the sensitivity of method including Deagle et al. (2007), Borovik et al. (2009), Tabesh et al. (2009), Mukherjee and Narasimhan (1996), Wu and Sage (2006), Wu and Sage (2007), Wu et al. (2010), Puust et al. (2006), and Boulos and Wood (1990). Pudar and Liggett (1992) formulated and solved the inverse problem using the areas of possible leak/burst events as the unknown parameter. Solution of the problem was achieved by minimising the sum of the squares of the differences between measured and modelled heads at selected points in the hydraulic model. This was achieved by using the Levenberg-Marquardt (LM) algorithm which allows solution of non-linear problems. The authors noted that this method is reliant on the selection of appropriate location and number of measurement points and is also highly sensitive to having a well calibrated hydraulic model with well-defined and accurate pipe roughness coefficients. The overdetermined problem (i.e. where more data points are available than the number of unknowns) offer the best results but this requires a significant number of instruments to be able to locate individual leak/burst events reliably. This would affect whether this approach could be adopted by water companies due to the financial constraints which are placed on their leakage management activities.

Mukherjee and Narasimhan (1996) modelled flow and pressure at the nodes in hydraulic models, both with and without leak/burst events, and uses the likelihood ratio test to judge how well the model fits the measured values of pressure or flow in the model. The model also assumes that the values for pressure and flow are not constant and will be subject to some random error, making the data more similar to the data which would be collected from a real network. The likelihood ratio test is used to detect, locate and size a leak/burst event given the measurements of nodal pressure and flow in both models and this was also tested on a small-scale laboratory rig which had seven branches. These results from the experimental study matched the simulated results well but it is not yet known whether this method is likely to be extrapolated to a WDS.

Deagle et al. (2007) applied the Synergee (now called Synergi) leak/burst event detection software package to three DMAs in order to test a method for calibrating hydraulic models using a genetic algorithm (GA) to improve the accuracy of the models for detecting leak/burst events. By modelling leak/burst events as pressure dependent and distributing them in order to match measured pressures within the network the location of leak/burst events and corresponding flow rates could be determined. This method required several months of flow meter data for a DMA which is usually available and one day of pressure data which is becoming increasingly common in the UK. The results of this study showed that this method is robust and can handle missing or spurious flow data very well but required a manual review of the data which is a drawback as this will increase the cost, time taken and subjectivity of the analysis of the results. By checking the predicted hotspots with the records of leak/burst event surveys in the area good agreement was found between the two, however several false positives were raised which in reality would be costly to investigate. This would suggest that the reliability of this methodology still needs to be improved to increase the confidence of the predictions.

Wu and Sage (2006) presented an approach for simultaneously calibrating a hydraulic model and detecting/localising leak/burst events using measured pressure and flow data from a real DMA. This approach required one logger per 200 properties to effectively calibrate the hydraulic model using a GA. An objective function was formulated to optimise the pipe roughnesses, demands at each node and operating status for each node in order to minimise the discrepancy between the measured pressures and the pressures predicted by the hydraulic model. The allocated demands were purely volumetric and not based upon the pressure in the network but were distributed between the network nodes evenly. A GA was used to find solutions to the objective function. This can produce different results for each GA run meaning that many runs of the GA are required in order to allow an average to be taken to counter this variability. The method was tested on both a simple simulated network and a real DMA from United Utilities' WDS which proved the methods ability to identify leakage hotspots. The main drawbacks of this method are the high volumes of data which are required in order to allow an accurate solution to be found and the accuracy of predicted nodal demands to allow effective calibration and hot spotting.

Wu and Sage (2007) further developed this method to include the emitter coefficient as part of the model formulation to estimate the size of leak/burst events in addition to the location. Pressure-dependent demands were modelled to ensure that the demand for each node was more realistic given the potential for low pressures due to the modelled leak/burst event scenarios. The predicted emitter coefficient at each node differed slightly for the evenly distributed leak/burst event scenarios generated using a simple hydraulic model as a benchmark. The variation was small in comparison to the overall magnitude of emitter coefficient. When a real DMA was considered the accuracy of the results depended on the number of allowable leakage nodes within the simulation although there is reasonable correlation between them. This methodology is limited when it comes to detecting smaller leak/burst events due to a lack of sensitivity and cannot distinguish between a single large leak/burst event and multiple smaller leak/burst events. Wu et al. (2010) built upon previous works by the authors to include up to 28 loggers in a test DMA in United Utilities WDS. The methodology is based upon calibrating the emitter locations and leak/burst event flows (by varying the emitter coefficients) throughout the hydraulic model to minimise the objective function. The demands in the previous studies were in no way related to the pressure at the nodes which has been addressed in this paper and tested on both a simulated network and a real DMA in United Utilities' WDS. This was improved by cross-checking the predicted leakage hotspots from the model with the leak repair records of the supply company. In a WDS typically the large, repairable leak/burst events can only be associated with a few dozen nodes within the network meaning that the optimisation problem can be greatly simplified to allow solution by using a GA.

The topics covered in this section have demonstrated how useful steady-state models are in identifying the potential sources of loss from water networks but have highlighted the limitations of the methods presented in the literature. The analyses have shown the ability to locate and size leak/burst events but are highly reliant on an accurate and up-to-date hydraulic model. Given the difficulties in obtaining all of the required information such as asset properties and operational status of the network elements there will always be doubts about the accuracy of the hydraulic model. This limits the effectiveness of this method for real-time applications as a lot of computations will need to be undertaken continuously on a number of data streams simultaneously to ensure a well calibrated model. This leads to further cost and makes the use of the steady-state methods less-appealing. Recently developed automatic online hydraulic modelling (Machell et al. 2010, Okeya et al. 2014) has been proposed as a potential solution to this problem by alleviating the manual work associated with hydraulic model calibration but these techniques are not developed to the point which renders them usable by water companies.

2.3.4. Transient Analysis-Based Techniques

Pressure transients (which will be truncated to “transients” from this point forward) arise in water systems as a result of sudden changes in pressure or flow. They are generally undesirable as they can lead to pipe failures, discoloration, noise and contamination and are usually prevented by good system design, surge vessels, surge anticipation valves, pressure-relief valves, air release valves and pump bypasses which are all means of preventing transients in the network (Boulos et al., 2005). The deliberate introduction of transients can be used to identify sources of loss from water networks. Transient techniques aim to make use of the response of a WDS system to transients to infer the presence and locations of leak/burst events. Pressure transients in WDS are generated by sudden changes in the operating conditions, such as pump shut-off or valve closure. Interpretation of the shape and speed of the generated transient can allow the detection and location of leaks. As transient waves pass through a pipe network they are shaped by the pipes they travel through and any fittings in the pipe network such as bends, tees, elbows, blockages, valves, reductions in internal diameter and changes in pipe condition. The basis of the transient based methods is that some property of the transient will be changed due to the interaction between the two and this can be measured and used to infer other information about the properties of the leak. Measuring transients requires high sampling frequencies in order to fully capture the detail of the events which often last only a short duration. When a transient interacts with a leak it acts to damp the initial wave and also causes a secondary wave to be reflected which can be used to both detect and locate leaks in the system if the reflected wave is significantly stronger than the reflected waves from the rest of the system. Often, however, the reflected wave will not be perceptible above the waves reflected from other pipe fittings and more information about the system is required to allow evaluation of the leak to take place. There are several classes of transient methods which will be reviewed in this section including inverse transient analysis, frequency domain techniques and direct transient methods.

2.3.4.1. Inverse Transient Analysis

Inverse transient analysis (ITA) allows solution of the inverse problem in the time domain in much the same way as the steady state-based analysis. The aim of ITA is to simultaneously calibrate the hydraulic model whilst also locating leaks in order to minimise the objective function. As per the steady-state inverse problem the aim is to minimise the discrepancy between the hydraulic parameters generated by the hydraulic model and the pressures from the network by altering the pipe roughnesses and the size and location of leaks. This process was broken down into three key stages by Kapelan et al. (2001); (i) the introduction of permissible transients into the system (ii) measurement/observation of operating characteristics (e.g. pressure/flow) at network locations (iii) calibration of the hydraulic model assuming the presence of orifice type leaks at certain locations in the network. The method was first conceived by Liggett and Chen (1994) and uses an iterative procedure to solve sequentially both the forward transient and inverse transient problems. The solved forward transient problem gives the relationship between the hydraulic parameters for a given iteration based upon a set of leak sizes, leak locations and pipe roughnesses. The inverse transient problem then modifies the pipe roughnesses and leak characteristics in order to alter the pressure at certain points to minimise the prescribed merit function by using the LM algorithm. The use of various solution methods such as GAs (Tang et al. 2001, Vitkovsky et al. 2000), LM (Kapelan et al. 2000) or hybrid GA-LM (Kapelan et al. 2004) to minimise the objective function has been found in the literature. A number of studies have been completed which utilise different methods to perform the calibration of the hydraulic model by either modifying the objective function to account for prior information (Kapelan et al. 2001, 2004) or finding more efficient methods of optimising the objective function.

2.3.4.2. Frequency Domain Response Analysis

There are a range of techniques which make use of frequency domain analyses to determine the frequency response of a WDS by converting time-domain data into frequency-domain data using Fourier Transforms. This characterises how certain frequencies of a transient are transmitted and absorbed by the pipeline and can give insights into the operating conditions of the WDS pipes. The frequency response function (FRF) of the network is governed by both the physical properties of the pipeline (such as roughness, diameter, bulk modulus and length) and operating conditions such as velocity and presence of connections and relates the magnitude and phase of the system output to the system input for different frequencies. Determination of the leak location is achieved by comparing the FRF of a pipeline with an induced leak with the FRF of a pipeline with no known leaks. In theory, the presence of leaks in pipelines can be revealed by additional resonant peaks occurring in the FRF (Mpesha et al. 2001).

Transients can be induced in WDSs by a number of means including the closure/oscillation of valves in pipelines (Ferrante and Brunone, 2003a, Ferrante and Brunone, 2003b, Gong et al. 2016, Mpesha et al. 2001) and the use of side-discharge valves. Once the transients have been generated the responses can be analysed to help size and locate leaks. The literature is replete with experimental studies which have been conducted using the FRF for leakage identification for a variety of scenarios including on single pipelines with single leaks and multiple leaks, parallel and branched pipelines with one leak and a simple reservoir-pipe-valve system with various leak sizes (Mpesha et al., 2001, Lee et al. (2002, 2005a, 2005b). Ferrante et al. (2016) conducted a study comparing the effects of randomly varying leak sizes, locations and numbers on the uncertainty of both time and frequency domain analyses. This showed that the local minima and maxima values of impedance are affected by the number of leaks for a given total outflow. Further studies have also included simulations of larger networks (Rashid et al. 2014) but no practical experiments have been conducted on real WDSs to test the effectiveness of the method.

Frequency-based methods present several benefits including a reduction in required measurement locations compared to ITA and that they can pinpoint leaks relatively efficiently (Datta and Sarkar, 2016). Also single and multiple leaks are readily identified (Colombo et al. 2009) which is a major benefit over ITA and there is the possibility of implementing the technique in near-real time so that leaks can be identified quickly although it is unlikely this could happen in the near-future due to performance limitations of the methods as they stand. Further studies are required in order to prove whether it could be economically viable to implement frequency response-based methods on a wider scale. Also problematic is the non-continuous nature of the analysis whereby transients are introduced and then the responses measured. The transient must be allowed to propagate through the system prior to any analysis taking place which delays leak identification and localisation. Also, the introduction of transients into the system will cause pressure fluctuations which could accelerate the growth of defects in pipe walls and contradicts the pressure reduction/network calming regimes which are being implemented by many utilities. Gong et al. (2013) stated that current FRF techniques are not suited to complex networks and linkages between transient events and discoloration of water in WDSs serves to discourage WCSs from utilising these techniques. Also, the introduction of transients into the system by using network valves will cause reduced downstream pressures and will potentially have an impact on service levels provided to customers. If side-discharge valves were to be fitted throughout the network this would increase costs above level required for pressure measurement alone. Several limitations were noted in a review by Colombo et al. (2009) including the difficulty in pinpointing leaks in a short pipe and the difficulty in detecting leaks at the quarter points of the conduit. As stated there are many barriers to applying frequency-based techniques to real WDSs which limit their potential for widespread application and the lack of examples to be found in the literature indicates this.

2.3.4.3. Direct Transient Analysis

To overcome the fact that an accurate hydraulic model is needed for ITA which can present difficulties in calibrating the model to allow accurate leak detection and location, methods have been developed which do not require an in-depth knowledge of the unknown parameters.

2.3.4.3.1. Time-Domain Reflectometry

When a transient passes a leak or other pipeline feature two effects are seen. Firstly, the initial wave will be damped slightly due to a loss of energy at the leak and a reflected wave will be generated which can be detected by pressure instruments deployed on the pipes. This class of techniques uses the arrival time of the reflected wave (hence time-domain reflectometry (TDR)) to the sensor to locate the leak by knowing the wave speed in the fluid. The size of the leak can be determined by the magnitude of the pressure drop when compared to an intact pipe. Pipe fittings, sediment/air accumulation and leakage all affect the generated pressure wave (Brunone, 1999) and experience is required in order to be able to differentiate particular phenomena from each other. Various studies have been conducted using TDR to locate leaks in single pipe systems for both water and wastewater systems. Brunone (1999), Covas and Ramos (1999) and Jönsson (2001) detailed applications of TDR. However, the method has only been applied to simple pipe systems and further development is required to prove this method on large scale, real WDSs. One of the only studies (Covas and Ramos, 1999) has been applied to a UK water network to validate the applicability of this approach. TDR methods are claimed to be simple and easy to implement due to the limited amount of equipment that is required but skilled personnel are required to interpret the complex data which is generated.

2.3.4.3.2. Advanced Reflection Techniques

Unlike TDR, advanced reflection techniques (ART) rely on more complicated analyses of the received pressure waves building on the TDR techniques reviewed in the previous sub-section. Beck et al. (2005) made use of cross-correlation of the reflections from singularities in the WDS to locate them. The authors demonstrated that the method presented high accuracy (5%) up to a distance of 95m even when the reflected waves were weak. The authors also claim that the number of sensors required to implement this method is significantly lower than suggested previously meaning that the application to real WDSs may be economical although this was only proved on a simple network of small-bore copper pipe with air being the contained fluid. Many reflections would be produced in an operating WDS as there is a much larger amount of network furniture such as valves or pipe fittings. Commonly used analyses for reflected transients are wavelet transforms and the impulse response function (IRF). Discrete wavelet transforms (DWTs) have been used by Stoianov et al. (2002) to detect leak reflections. DWTs are used for detecting discontinuities in a pressure signal which are affected by the shape of the injected transient (Colombo, 2009).

The impulse response function (IRF) is the relationship between the system response to a given input transient event. An impulse response is only affected by the physical properties of the system and is not affected by the shape or magnitude of the induced transient. Vitkovský et al. (2003) found that IRF can be used to convert system reflections to sharp edge impulses and hence they can be more easily detected. Lee et al. (2007) performed an experimental trial to obtain the frequency response for a pipe between two pressurised tanks. The authors noted that rapid changes in operating states are required to generate transients which can catch pipeline features in enough detail to be of practical use. The approach is still highly dependent on a good approximation of wave speed. Limitations exist when using these techniques. Firstly, the methods that rely on comparing the leak free operating state with the leaking operating state cannot discover any information about the leaks which were initiated prior to the start of the analysis period (i.e. before the pressure information started being recorded). Also detailed system knowledge is required in order to create an accurate leak-free model with which to compare the obtained pressure responses and errors can be caused by incorrect system information.

2.3.4.3.3. Transient Damping Techniques

Another class of transient techniques which have been developed are those utilising the damping of transients to measure leak sizes and locations. The studies found in the literature utilising this technique include Wang et al. (2002), Covas et al. (2005), Lee et al. (2005a, 2005b) and Nixon and Ghidaoui (2007). In order to determine the amount of transient damping the Fourier response of the system is assessed prior to arrival and after interacting with the leak. Typically, a Fourier analysis will be undertaken to see how each of the frequencies responds to the leaks in the system. It has been shown in the literature that leaks down to 0.1% of the cross-sectional area of the pipeline can be detected (Colombo et al. 2009) although this only applies to simple systems and conditions which are not likely to be encountered in real WDSs. Frequency-based techniques are highly sensitive to background noise which makes application to real WDSs challenging. Lee et al. (2005a) also highlighted that two leaks which are equidistant from the monitoring location will be registered as a single leak, and whilst this scenario is unlikely this could be potentially a major limitation.

2.3.5. Negative Pressure Wave Techniques

Negative pressure wave (NPW) techniques are a class of techniques aiming to exploit the generation of a transient upon initiation of a leak event in a WDS. Unlike other transient based techniques, the method is aiming to detect transients produced by the WDS rather than those imposed on the system. When a leak event starts in a WDS a low-pressure wave is generated which can be detected by pressure measurement devices. The pressure wave travels up and downstream from the leak at an approximately constant speed which is much higher than the bulk flow in the pipe allowing localisation by calculating the time of arrival if there are multiple sensors installed in the WDS. Equations for calculating the arrival time at each of the sensors is presented by Tian et al. (2012). There are a number of challenges to using this approach which have thus far limited its application to water networks including; data quality, sensitivity to system dynamics and a high false alarm rate.

Localisation of leaks cannot be achieved by pressure analysis alone (Sheltami et al. 2016) making the installation of flow meters necessary in order to utilise the flow balance method. By studying the normal transients generated in the pipeline this allows a threshold to be set which can determine whether a leak is present or not. Using statistical process control (SPC - Oakland, 2003) to create an adaptive threshold is useful to cope with a variety of working conditions of the pipeline. This can help to overcome the high rate of false positive leak identifications which have been found when using traditional pressure analysis techniques, such as the cumulative sum (CUSUM) algorithm. The magnitude of the leak event can be calculated from the measured flow rates before and after the leak was initiated. De Joode and Hoffman (2011) presented the ATMOS Wave method which showed good reliability at detecting leaks when tested on over 100 leak trials in a variety of network types (water, gas, jet fuel). One drawback of this method was the presence of blind spots near to the sensor which will cause loss of leak identification. To overcome this, two sensors are required at each measurement location, increasing the instrumentation costs.

There are records of a number of experimental studies to validate the NPW approach in laboratory conditions for gas pipelines (Shuqing et al. 2009, Hou et al. 2013) and water pipelines (Misiunas et al. 2005, Silva et al. 1996). Field testing has also been conducted (Srirangarajan et al. 2010a, 2010b, Misiunas et al. 2005, Hu et al. 2011, De Joode and Hoffman, 2011) for a mixture of different pipe sizes and fluid types. It must be noted however, that the studies have only been conducted on single pipelines or small network configurations with a few pipes and relatively high concentrations of loggers. Even under these conditions the performance of the NPW method shows some advantages including; fast detection time, accurate leak location and high sensitivity (De Joode and Hoffman, 2011). The sensors in this study were placed at either end of a single transmission pipeline in each but one case meaning that this method cannot easily be transferred to a network with loops which are nearly always present in a DMA.

2.3.6. Data-Driven Techniques

Data-driven techniques for detecting leaks and bursts in WDSs are becoming increasingly popular. This has mainly been influenced by the reduced cost of installing and maintaining instrumentation on the network (Romano et al. 2016). The increased quantity of data which is generated means that traditional approaches must be replaced with something more robust and efficient in order to be able to deal with spurious or missing data and to handle the high data volumes in a timely manner. Data-driven techniques have been used for the last 20 years to analyse pressure, flow and water quality data taken from hydraulic models, experimental laboratory apparatus and real WDSs to give insights into the operational status of the networks and infer the presence and location of leaks. The earliest studies by Skipworth et al. (1999), and Mounce et al. (2002, 2003) made use of artificial neural networks (ANNs) to analyse flow or pressure data (or both) to give insight into the operating state of a WDS and monitor when an abnormal operating condition has been reached. The flow data in these studies was collected once per minute and up to 10 portable pressure loggers were used within one DMA with sampling frequencies of one reading every 12 seconds. Mounce et al. (2003) and Khan et al. (2006) also made use of opacity (turbidity) sensors to correlate the increase in turbidity with the increase in flow to allow localisation of the leak.

The rest of this section will classify the data-driven approaches according to the methodology used and provide insights into the benefits and limitations of each of the classifications. It is also worth stating that a number of data sources can be used in order to develop a leak detection methodology. The data can be collected from pressure sensors deployed in a real world WDS. This gives the best performance in terms of the methodology being able to handle real network conditions and being applicable by the WSCs, although the presence of leaks/bursts in the data set cannot always be guaranteed. Secondly, data can also be generated using a hydraulic model which can allow rapid generation of vast quantities of data but this is not representative of the random behaviour which is seen in real WDS. This can lead to difficulties when transferring a methodology which has been developed using simulated data (using a hydraulic model) to a real WDS. Lastly, data can also be collected from rigs in laboratory conditions. These networks will not be very complex (often composing a single pipe or loop) but conditions can be controlled very precisely to allow the relationship between variables to be established with great certainty. Where these approaches are used in the literature will be highlighted in the following sections. The techniques are divided as per the review by Wu and Liu (2017).

2.3.6.1. Classification Techniques

Classification techniques aim to distinguish between normal and abnormal operating conditions by identifying the features of the data which represent each type of condition and matching the newly acquired data to the class which best fits its features. The features of each data set (e.g. normal and abnormal) are contained within the data and can be used to distinguish subsequent data sets by training a classifier to recognise the features. Historically ANNs have been used extensively due to their classification ability when presented with a representative set of training examples containing known leaks and bursts (Caputo and Pelagagge, 2003, Romano et al. 2014a). ANNs are based upon the working mechanism of the human brain and have long been recognised for their ability to recognise patterns in data when supplied with a sample data set (Bishop, 1995). ANNs are able to transform a set of input variables to output variables by adjustment of a set of weights. The weights are adjusted to minimise the difference between required output and the achieved output of the ANN during training by minimising the error term (e.g. mean squared error, mean absolute error, and mean absolute percentage error). There are several advantages of ANNs. Firstly, they overcome the need of having a detailed understanding of a process as thy are able to learn how a process behaves using data collected from the process itself. They are capable of handling noisy/incomplete data and can be periodically retrained to cope with changes in the process. One of the major problems when using ANNs is the dependence on the quality and amount of training data which has a big impact on the reliability and accuracy of the results obtained. Overfitting, which can occur when too much training of an ANN takes place and the ANN cannot be generalised to a different data set such as the test/validation data set. Therefore, overfitting can lead to poor performance of an ANN and selection of the number of training data is critical. Mounce and Machell (2006) utilised both static and time delay ANNs to classify flow data for leak/burst event detection. The results of this study show that the time-delay neural network (TDNN) performed best with a total delay length of 40 minutes. The mean-squared error (MSE) was shown to be an order of magnitude lower than for the best performing static neural network. It must be noted that to successfully train ANNs an extensive data set containing engineered events is required, which is not always possible. Aksela et al. (2009) used a self-organising map (SOM) which is a type of unsupervised graphical ANN depicting a discretised representation of the input data. In the study the SOM was used to detect when leak/burst events had occurred based upon flow meter data and mains repair records collected from a Finnish water company. The hydraulic impact of the leak/burst event (measured using the leak function) indicates the location of the leak/burst event from the flow meter.

Support Vector Machines (SVMs) have also been proposed as a way of classifying data into both normal and abnormal categories as they behave in a similar way to ANNs. They are trained on data in the same way as ANNs but they go about classification between data groups in a different way. ANNs rely on back-propagation and gradient descent to minimise the error function whereas SVMs will solve a constrained quadratic optimisation problem. Mashford et al. (2009) used SVMs for classification of simulated data from an EPANET model in order to detect and locate leak/burst events. A comparison of ANNs and SVMs for intruder detection (which can be considered a binary classification anomaly detection problem along with leak/burst detection in WDSs) by Mukkumala et al. (2002) showed that the SVM performed better with respect to accuracy of classification and a reduction in training time. No examples of SVMs being used to analyse data collected from a real WDS have been found in the literature although there is no obvious reason why this should be the case. Generally, ANNs are favoured for this purpose.

2.3.6.2. Prediction-Classification Techniques

Another class of techniques which have been identified is a combination of predicting future values of one or more hydraulic parameters and then classifying the status of the WDS based upon the discrepancy between the predicted value of the parameter and the value obtained from the sensors deployed in the WDS. Several studies utilising both prediction and classification can be found in the literature including; Mounce et al. (2002, 2003, 2010, 2011), Romano et al. (2010, 2014a, 2014b) and Bakker et al. (2013, 2014) all utilised prediction-classification techniques for either leak/burst event detection, localisation or both. The basis of the prediction-classification techniques is that no examples of anomalies are required in the collected data as these are identified by a data-driven model. This makes the classification of data much simpler but there are difficulties applying these models to real WDS which limits their potential impact.

Caputo and Pelagagge (2002) proposed an early study which utilised a combination of hydraulic modelling to simulate the effect of leak/burst events at nodes in the model and ANNs to link these with potential causes. Caputo and Pelagagge (2003) built upon this methodology by applying it to data obtained from a model of a district heating system. Leak/burst events between 2%-10% were simulated at a range of nodes and a 2-layer ANN was used to locate the leak/burst events. The reported precision was that for 28 out of the 30 branches in the model it was less than 40m. The use of ANNs has been proposed more recently by Mounce et al. (2007) to predict the flow at a district inlet meter. A back-propagation mixture density network (MDN) was used to predict the flow and this was subsequently compared to the observed flow to leave a residual which can then be classified using a fuzzy inference technique, called the Mamdani method (Mamdani and Assilian, 1975). This will output a fuzzy logic value for the presence of a leak/burst event at a given confidence interval for a particular time window. Once a burst has been identified by the fuzzy inference system the flow rate was also estimated from the difference between the predicted and actual flow rate. Mounce et al. (2010) presented a near real-time application of the same leak/burst event detection system. These studies included the analysis of data from 146 DMAs over the course of 2 months. The main problem encountered was the relatively high false alarm rate of 18% which could lead to a lot of wasted resources investigating them. Several alarms were also caused by data acquisition issues and abnormal network events. This leads to a lack of confidence in the causes of a particular alarm and hence can reduce the appeal of the system to water companies. No comparison was carried out between using the automatic AI solution and the manual nightline analysis that is routinely undertaken by the water companies although the authors have highlighted this as a future target.

Romano et al. (2010) used ANNs as one of the stages within their leak/burst event detection system to predict future pressure and flow values one-time step ahead. Discrepancies between the predicted and the observed values were treated as indicative of problems within the WDS. The discrepancies were analysed using statistical process control and a Bayesian inference system (BIS) calculate the probability of a leak/burst event being present within the DMA being analysed. This methodology was improved by Romano et al. (2011) to include a geostatistical analysis to interpolate between the location of the deployed sensors and give a predicted location for a burst in the DMA. A comparison of several geostatistical techniques (e.g. ordinary kriging (OK), ordinary cokriging (OC), local polynomial (LP) and inverse distance weighted (IDW)) was undertaken by Romano et al. (2013) in the context of localising leaks in a DMA based upon real data. This showed that OC allowed for the smallest area predictions for potential bursts (i.e. the greatest accuracy). Further alterations allowed for the automated analysis of the large amounts of data required to perform the previous analyses (Romano et al. 2014a) and included further information about the operating status of network assets such as valves to be included (Romano et al. 2014b). These methods have all shown good ability to not only detect but also localise leaks/burst events within real WDS which represents a significant advancement in the field.

One problem which has been noted is the seemingly random fluctuation of pressure and flow within a real WDS which makes prediction and analysis particularly difficult. Techniques have been utilised which allow the “noisy” characteristics of the signal to be removed. Several studies have utilised Kalman filters (Ye and Fenner, 2011, Okeya et al. 2014, Jung and Lansey (2014, 2015)) which allows prediction of the flow or pressure at future time steps. The difference between the output of the filter and the observed flow rate gives the predicted amount of water lost in the DMA and the filtered pressure signals can be used to provide confirmation of the flow analysis if required. This method was applied to flow data from UK DMAs and the results indicated that the flow data is more sensitive to bursts than the measured pressure hence this method is limited, as it cannot provide any information regarding the location of the leak or burst due to this.

Bakker et al. (2014) devised a method for predicting the water and pressure sensor values within a DMA 12 hours in advance (in 15-minute time steps). The model is capable of handling the different days of the week and can also account for exceptional demand days (such as public holidays and school holidays) to produce flow and pressure predictions. The results stated that for the largest of three supply areas (130,920 properties) only the largest 7.8% of bursts were detected (hence why the recommended DMA size is 2500 properties or less) as opposed to 50% for the smallest supply area (650 properties). The authors stated that the method was effective at reducing the time taken to detect the leak but that the time to locate the leak was still problematic. A location method should be implemented to reduce the overall run time and hence impact of a leak or burst on customers and the water companies.

2.3.6.3. Statistical Techniques

The measurement of pressure and from sensors located in a DMA allows certain characteristics of the signal to be monitored only in relation to itself (i.e. a statistical measure). This forgoes the need of creating a detailed hydraulic model which can often be difficult to calibrate and can give insights into the presence and locations of leaks and bursts at or near the sensor locations.

The basis of the statistical methods is to compute variation of some parameters (e.g. mean or standard deviation) over time. Deviation above a previously determined level will raise an alarm to preclude intervention or further analysis from the operators of the WSC. SPC is a method which was previously used to control the quality of products manufactured on production lines and there are a number of variations in the literature including; cumulative sum (CUSUM) – Bakker et al. (2014), Misiunas et al. (2005), exponentially weighted moving average (EWMA) – Kim et al. (2016) and Shewhart individual control charts – Romano et al. (2016). CUSUM estimates the severity of the change in a variable over time and once this exceeds a pre-defined threshold in order to detect a step in the data which would be indicative of a change in operation or a potential leak/burst. There are difficulties in applying this algorithm however as two parameters (the detection threshold and the change magnitude) must be correctly set to tune the algorithm and ensure it operates as desired. This can be problematic and requires the mean detection delay to be set which can impact upon how quickly the algorithm can detect leaks or bursts, and hence its efficacy for online applications. EWMA is a moving mean calculated over some time window whereby the mean is weighted so that the older data in the time window have less impact on the mean. The weighting decreases exponentially as the age increases, hence the name. Kim et al. (2016) utilised the cumulative integral on a data set which was filtered using a Kalman filter and a floor function was applied to set a threshold of 10% of the mean pressure to identify when a leak or burst has occurred. Control charts set the alarm thresholds at multiples of the standard deviation either side of the mean. Various sets of rules such as the Western Electric Rules, Wheeler Rules and Nelson Rules (Oakland, 2003) have been proposed to reduce variability of quality in manufacturing and could be applied to the pressure and flow signals obtained from a WDS to monitor variations for leak and burst detection purposes.

Principal component analysis (PCA) is a statistical technique which converts a set of loosely correlated variables into a set of linearly uncorrelated variables which are termed principal components. The procedure has been applied for burst detection when applied to DMA flow meter data by Palau et al. (2004, 2012). The flow data was divided into classes representing the different times of day (night, morning and afternoon) to allow different statistical models to be built for each. Principal component analysis has also been used for the detection of bursts in WDS by Palau et al. (2004, 2012). The daily inflow data was collected over a 6-month period on an hourly basis and any days where known bursts or illegal consumptions were removed to ensure the model was sensitive to detecting these. The data set was then processed (mean-centered and scaled) to normalise it, divided into weekdays and weekends and 4 models were built (one for each time of day and one for each day as a whole). The number of principal components which described each data set were extracted and validated by identifying and removing severe and moderate anomalous readings before control charts (e.g. T^2 Hotelling and Distance to model) were established to allow failures to be detected in the data sets. This method is computationally efficient as the dimensionality of the space if severely reduced and bursts down to 5% of the total demand could be detected with up to 95% confidence. Palau et al. (2012) also presented a case study of PCA applied to real DMA data indicating its potential to be applied by water companies. As the flow from the whole day is analysed the speed of burst detection is likely to be a matter of hours, which could greatly reduce the runtime especially for small bursts. This method is highly dependent upon the quality of the acquired data so pre-processing is essential to maintain sensitivity, which is crucial for any burst detection methodology.

Van Thienen (2013) presented a methodology for analysing flow meter data by using the comparison of flow pattern distributions (CFPD). The method allows for consistent and inconsistent changes in the flow to be identified which can distinguish between changes caused by population characteristics (consistent changes) and new usage by large consumers or changes in leakage in the distribution system (inconsistent changes). The method is a graphical representation of the comparison of the data obtained on different weeks. The data are divided up (either by day or week) and a small section of the flow data is extracted for the same time period in the given interval. These data sets are plotted against each other and a line of best fit is plotted through it. The magnitude of the slope and intercept and then organised into a matrix and coloured according to their magnitudes. This allows the user to visualise where large changes have taken place and where remedial action is required. Further studies have applied improvements to this methodology to DMA flow meter data (Van Thienen (2013) and Van Thienen and Montiel, (2014)). One important limitation of this method is that the location of leaks cannot be established and the lengthy search times will still be problematic.

Evolutionary polynomial regression (EPR) was proposed by Giustolisi and Savic (2006) as a “grey-box” mathematical modelling technique which incorporates prior knowledge of the physical system modelled before, during and after calibration of the EPR model. A small amount of data from the physical system (i.e. pressure and flow) is required to allow the model to be calibrated to conserve accuracy in the final model. The general model is formulated as the sum of a series of polynomial terms where the coefficients and the exponents are to be determined by an evolutionary search algorithm to give the best fit to the measured data points. To minimise the search space a list of exponents can be pre-selected to be solely used by the algorithm. The model is thought to match the data sufficiently when some previously chosen error function is reduced to an acceptable level. In Giustolisi and Savic (2006) the sum of squared errors was used for selection of the most appropriate model. The methodology was applied finding the Colebrook-White equation for a range of Reynolds numbers and pipe roughness values.

Laucelli et al. (2016) applied EPR to flow and pressure data taken from a real DMA. The data was taken from the inlet, outlet and a critical pressure monitoring point within the DMA and was not pre-processed to avoid the removal of various features which could lead to overfitting of the EPR model. The selection of models was formulated to penalise the number of explanatory variables and polynomial terms. This methodology also incorporated a mechanism for identifying deviations from normal operation by applying a normal probability distribution to the operating status of the network and selecting a 95% cut-off threshold between normal operation and anomalies. This procedure was effective in identifying a number of incidents which was confirmed by the records of customer contacts and interventions made by the water company. One problem with this methodology is that small leaks can easily be masked by larger ones and will not become apparent until the larger leak is fixed. This method, in isolation, cannot provide any information regarding the location of leaks which is a drawback.

2.3.7. Summary of Leak/Burst Detection and Localisation Techniques

Currently a combination of hydraulic and hardware-based techniques is used throughout the UK water industry. Some hydraulic techniques, such as the water audit, allow the assessment of entire WDS, or section thereof, and enable the targeting of resources to improve the performance of the WDS in terms of water loss. However, water audits do not provide sufficient information regarding individual leak/burst events. Water audits cannot be carried out in real-time or near real-time and as such there will typically be delays between leak/burst events occurring and a water company being aware of them. Given the changing landscape within the regulation of the UK water industry it is fair to say that reactive techniques, such as the hydraulic techniques currently used by water companies, will not be the best techniques going forward when a more proactive approach to the management of water distribution systems is required. In recent times, automatic near real-time techniques have been developed and implemented to detect leak/burst events. This represents the current state-of-the-art in leak/burst detection and localisation in the UK water industry. Hardware-based techniques, on the other hand, can determine the presence and exact location of individual leak/burst events so that interventions can be made to restore the performance of the WDS in a short time. Hardware-based techniques are resource intensive meaning that they must be planned to prioritise the most damaging leak/burst events. By combining hardware-based techniques with hydraulic techniques so of the drawbacks of both can be minimised to improve the overall efficiency of leak/burst detection and localisation.

It is clear that the problem of detecting and localising leak/burst events in water distribution systems is still an open problem which has not been solved yet and room for improvement remains. A wide range of techniques have been developed, based on networks of sensors deployed throughout a WDS, and some even applied to WDSs but industry wide adoption of these techniques by water companies has still not been seen. Generally, the goal of developing new leak/burst detection and localisation techniques is to have them adopted and used as part of their daily operations by water companies. Water companies have historically been very conservative with regards to the adoption of new technologies to reduce their exposure to risk which means that the developed techniques must be well-proven prior to being adopted. In addition to this there must be a very concrete benefit or combination of benefits of new techniques to be adopted by water companies. This could be, for example, improved performance or reduced costs. The reason that many techniques are not adopted is that they all exhibit at least one major drawback which makes them unappealing to water companies around the world. A related factor is that many techniques purport to provide potential benefits but the level of confidence around this benefit is not high enough for water companies to invest in them. Of the techniques which have been reviewed there are two standout examples of leak detection techniques which have been adopted and commercialised in the water industry. These are Romano et al. (2014a) and Mounce et al. (2010). The ANN based near real time leak/burst detection system developed by Romano et al. (2014a) has been implemented operationally by United Utilities in the UK. The leak detection system developed by Mounce et al. (2010) has been commercialised by Ovarro (Ovarro, 2020) for use across multiple sectors and is called “Datective”. The same cannot be said for localisation at the current time which is likely due to the increased number of instruments required for localisation versus detection and the relatively small number of techniques which have been developed when compared to leak/burst detection. Directly related to this is the fact that trialling new technologies using data collected from real water distribution systems has, until quite recently, been very difficult. This was linked to the desire of water companies to not interrupt their day-to-day operations and also to avoid the inherent risks associated with implementing new technologies. This means that the performance of any techniques was typically only evaluated using either synthetic data or a small number of real leak/burst events. This outlook has changed in the last few years and more techniques than ever are being developed and tested using real data. The latest trends for leak/burst localisation pressure sensor deployment show that sensor densities of fewer than five pressure sensors per DMA are commonly deployed (Soldevila et al. (2017, 2019), Perez et al. (2011), Cuguero-Escofet et al. (2015). Examples of widescale deployment of multiple additional pressure sensors throughout multiple DMAs are rare although some do exist, such as Romano et al. (2016). On average Romano et al. (2016) deployed 5 sensors per DMA in 17 DMAs which is the largest deployment of its type for leak/burst localisation using pressure. Even in this case, the deployment of pressure sensors was for the purpose of testing a leak/burst localisation technique and not deployed as permanent assets. In the UK, there are some large case studies of deployment of multiple different types of sensor, usually for multiple purposes simultaneously. For example, the NEPTUNE project (Savic et al., 2013) featured temporary deployment of pressure sensors (up to 13 per DMA) across a section of Yorkshire Water's WDS. The pressure data, in combination with the flow data collected from each DMA inlet, was used for leak/burst detection and incident response. The biggest example of sensor deployment which can be found anywhere in the literature was by Affinity Water, who are based in several regions across the UK. Approximately 20,000 acoustic sensors were permanently deployed throughout an entire WDS (Affinity Water, 2018) in order to continuously monitor leak/burst noise and quickly localise new leak/burst events. On the international stage the story is broadly similar. The Smart Water for Europe project (Vitens et al., 2018) spanned four European countries with a test site in each. Of the four sites, three used a combination of additional pressure, water quality and flow monitoring as a means to detect and localise leak/burst events. In total more than 4000 sensors were deployed across the four test sites on a temporary basis. Based upon these examples, there is a clear need for greater permanent deployment of sensors in WDSs so that water companies can proactively monitor them on a long-term basis.

Those techniques which have been adopted by water companies are typically aimed towards detecting leak/burst events rather than localising them. However, there are a range of potential benefits in using an automatic, (near) real-time system for localising leak/burst events rather than just for detection. Given that the resources which are required for leak/burst detection, location and repair is currently high in a normal WDS, there is still the potential for improving the overall efficiency of the process. The first of these is to use automatic detection and localisation using existing instrumentation aided by additional pressure instrumentation. This can direct field teams to the approximate location of a leak/burst event so that the overall time for finding and fixing the leak is minimised. There are several other benefits of using this approach. Firstly, augmenting existing leak/burst localisation procedures makes use of the existing knowledge of water company personnel and hardware which is already owned by the water company. Given that most water company hardware will have been purchased for use over multiple asset management periods then maximising the useful life of this equipment can allow water companies to make savings which can be used for other purposes. On top of this, using additional pressure instrumentation can fulfil other business needs for a water company. Examples of other needs which can be served by pressure instruments are hydraulic model calibration or more accurately reporting low-pressure events (for regulatory reporting).

Table 2-1 *Performance comparison for hardware-based leak/burst detection and localisation techniques*

Technique	Case Studies	Cost	Accuracy	Event Sizes	Intermittent/Permanent
Listening Devices	Industry-Wide	Low	Low	Large	Intermittent
Leak-Noise Correlators	Industry-Wide	High	High	Small	Intermittent
Noise Loggers	Real	High	Medium	Small	Both
Tracer Gas	Laboratory	High	High	Small	Intermittent
Thermographic	Laboratory	High	Medium	Large	Intermittent
Ground Penetrating Radar	Laboratory	High	Medium	Large	Intermittent
Pipeline Inspection Gauge	Real	High	High	Small	Intermittent

Table 2-2 *Performance comparison for hydraulic leak/burst detection and localisation techniques*

Technique	Case Studies	Cost	Accuracy	Event Sizes	Intermittent/Permanent
Water Audit	Industry-wide	Medium	Low	N/A	Intermittent
Step Tests	Industry-wide	Medium	Low	Large	Intermittent
Steady-State Analysis	Real	High	Medium	Small	Intermittent
Transient	Laboratory	High	High	Small	Intermittent
Negative Pressure Wave	Laboratory	High	High	Small	Intermittent
Data-Driven	Real	Medium	High	Small	Both

2.4. Sensor Placement Techniques

2.4.1. Introduction

Sensor placement techniques are used for many purposes within the analysis of water distribution systems. Having said this there are many features which are shared between techniques even when they were not developed with the same goal in mind. This section will present a review of the available techniques which have been developed for placing sensors and group them according to their purpose. Special attention will be paid to those techniques which were developed for the purpose of placing sensors for leak/burst localisation to identify gaps in the literature and to determine where improvements are needed to increase the performance of the techniques and to increase the rate of adoption of these techniques by water companies. This last point is particularly important and, as will be demonstrated, there are still many techniques which have been developed which have never been tested outside of highly idealised cases using either hydraulic models or synthetic data sets which do not pose the same difficulties as real data.

The differences between the goals of different sensor placement techniques and the number of studies mean that there is no overarching review of all sensor placement techniques. The types of sensor placement techniques can be divided into the following areas:

- Leak/burst event detection and localisation
- Hydraulic model calibration
- Contamination event detection and localisation
- Water quality model calibration

Each of these areas has been the subject of differing amounts of research. Review papers relating to the areas of placing pressure and flow sensors for leak/burst event detection and localisation (Romano, 2019) and placing water quality sensors (using either a single or multiple variables) for contamination event detection (Hart and Murray, 2010) are available. For calibrating hydraulic models and water quality models no standalone review papers, dedicated solely to the sensor placement aspect, are available. However, reviews conducted as part of individual studies are available and will be used here instead.

It is worth highlighting here that there is a coupling between the types of sensors which are favoured and the types of events which are being detected and localised/models being calibrated. This is driven by the typical (and normally most obvious) effects which are seen for different types of events. For leak/burst event detection/hydraulic model calibration pressure and flow sensors are used because the primary effects of these events are seen as changes in pressure and flow. For contamination event detection and localisation/water quality hydraulic model calibration water quality sensors are normally used. When water quality sensors are used a range of different variables can be measured which are defined by the goals of the detection/localisation system or the model calibration.

2.4.2. Leak/Burst Events

2.4.2.1. Overview

Sensor placement techniques for leak/burst detection and localisation are used to deploy sensors in a configuration which allows leak/burst localisation or detection techniques to achieve the goal(s) for which they were developed. Different goals are often used and are driven by several aspects. Firstly, the strategic business aims of water companies will vary depending upon the regulations they are subjected to and the resources that they have available to achieve those strategic aims. In the UK, for example, a company may have agreed to reduce its leakage by a certain amount or to reduce the duration of supply interruptions. These goals can be achieved in different ways and, therefore, a sensor placement technique will need to consider these goals when choosing where to place sensors. The difficulty is that these aims are not often considered by academic researchers who are developing the systems which will later be implemented by the wider water industry.

An important point is that the widespread monitoring for both pressure and flow in many water distribution systems across the world was initially driven by the implementation of DMAs (Farley and Trow, 2003). The main result of implementing DMAs was to allow for better quantification of parts of the water balance (Liemberger and Farley, 2004). Leak/burst detection and localisation techniques aim to build upon this by allowing a more proactive approach to managing leak/burst events which can contribute to lower losses from a WDS. The majority of sensor placement techniques are set against the backdrop of DMAs already having been implemented in the WDS which are being considered. Over the past few years, the UK water industry is undergoing somewhat of a revolution as sensor technology improves and becomes cheaper. This situation combined with increased regulatory pressure to improve performance and efficiency in many areas related to the operation and management of WDS means that there is a higher level of interest in (and rate of installing) pressure and flow sensors within WDS than ever before. Sensor placement is important as it allows the installation process to be targeted towards the strategic aims of a water company. On top of this, numerous techniques have been developed which can determine the optimal sensor configurations which can allow water companies to achieve their strategic aims and regulatory targets. The next two sections will cover the techniques which have been developed and extract out some of the common features between them.

A review by Romano (2019) was the first attempt to aggregate and compare the available techniques for deploying flow and pressure sensors in water distribution systems for the purposes of leak/burst detection and localisation. Some key issues related to developing sensor placement techniques for leak/burst detection and localisation were identified which included accounting for uncertainty in both hydraulic models and sensor data, accounting for multiple leak/burst event sizes, accounting for failure of sensors and communications systems, accounting for risk and the wider need to include flow instrumentation. The key finding was that none of the techniques accounted for all of these aspects. A second important issue which was identified by Romano (2019) was that there is a very limited amount of application of the techniques to real leak/burst events in WDS. This is a key area for improvement as the ultimate litmus test for any sensor placement technique is whether it can identify configurations which perform well on real data as opposed to just hydraulic model simulated data.

In Figure 2-1 several steps that are common to sensor placement techniques for leak/burst detection and location have been identified. The next two subsections will outline the key features of the sensor placement techniques which have been found in the literature, in relation to the steps in Figure 2-1, and compare and discuss the performance of these techniques. The detection and localisation techniques have been divided here. Where a sensor placement technique is aimed at achieving both detection and localisation then this is included in the localisation sub-section.

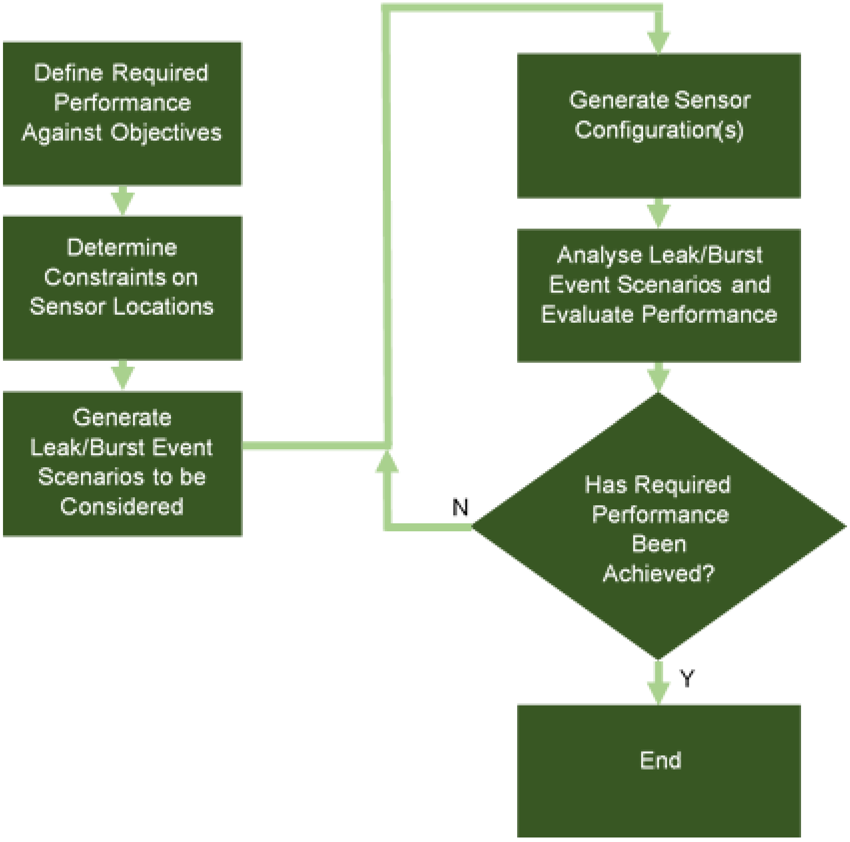


Figure 2-1 *Schematic of a generic sensor placement technique for leak/burst detection and localisation*

2.4.2.2. Detection

Detection, as defined by Puust et al. (2010), is the process of a water company becoming aware of the presence of a leak/burst event in their WDS. There are many sensor placement techniques which aim to place sensors for the purpose of detecting leak/burst events. Most sensor placement techniques are considering the placement of pressure sensors because many WDS have flow monitoring due to the implementation of DMAs. Some techniques also consider additional flow monitoring, however, and where this is used it will be highlighted.

Firstly, several different objectives have been used to evaluate the quality of sensor configurations. One objective for all of the techniques which are reviewed here is the number of sensors which are used. This is logical as the number of sensors is one of the factors which most impacts upon the level of detection performance which can be achieved. Also, from a water company perspective, the number of sensors is closely related to the cost of a sensor configuration and must be considered against the performance to select a suitable monitoring strategy. The number of sensors is treated in different ways depending upon the sensor placement technique which is being used. Farley et al. (2008) considered the installation of exactly 2 additional pressure sensors (assuming that the inlet was also monitored for pressure) within a DMA. To validate the methodology using real data Farley et al. (2010) considered several different instrument locations individually and compared the performance using modelled data and real data. This demonstrated good agreement between both data types highlighting the potential performance ability of using the methodology operationally. Forconi et al. (2017) used a similar approach which used the max-sum (Bush and Uber, 1998) to rank individual pressure sensor locations. An extension of these techniques was to consider a fixed number of pressure sensors but repeatedly run the analysis to examine the performance using different numbers of pressure sensors. Wu and Song (2012) considered between 1 and 5 pressure sensors and Puleo et al. (2018) considered between 2 and 5 sensors. This allows the solution space to be narrowed considerably for each run but does take more time because individual runs must be performed. However, the most important issue is that the trade-off between performance and the number of sensors can only be apparent when different numbers of sensors are considered. Lastly, Raei et al. (2019) considered the multi-objective nature of the problem and directly optimised the number of pressure sensors and the detection performance. This is seen as the best approach as multiple optimal solutions (with different numbers of sensors) can be determined simultaneously using a multi-objective optimisation technique, such as the NSGA-II GA (Deb et al. 2002).

Two techniques were found which considered the placement of both pressure and flow sensors. Hagos et al. (2016) determined the detection performance for a fixed number of sensors and ran the sensor placement technique repeatedly to demonstrate the cost/performance trade-off. However, they did not account for the different costs of flow and pressure instruments which is important as in reality there is a big disparity between the two which limits the desirability for deploying multiple flow instruments in a DMA. Raei et al. (2018) proposed a two-step strategy which first determined the optimal configuration of flow sensors using a multi-objective approach. The pressure sensor locations were then optimised subsequently considering the determined configurations of flow sensors. The number of flow sensors considered was large (a minimum of 5 flow sensors) however the model used to demonstrate the technique was as large as a city so the number of meters was considered realistic.

A different approach used by Huang et al. (2012) did not consider the number of sensors directly but considered the structure of a SOM (Kohonen, 1982). This is problematic as the solutions which were determined by the technique had similar numbers of pressure sensors and did not vary in performance significantly. This meant that the trade-off between the performance and cost was not fully demonstrated and is of little practical applicability.

The second objectives, where used, were most commonly performance related. A range of different performance measures were used to evaluate the sensor configurations. The most common was to maximise the number of events which were detected. Farley et al. (2010) achieved this by using a binarised sensitivity matrix, where the values were converted to either 0 or 1 depending upon their magnitude in relation to an arbitrary threshold. The sensors with the highest score were chosen as the best locations. Previously, Farley et al. (2008) had used the raw sensitivities, without binarisation, in the same way. This method of determining the most sensitive locations was adopted from Bush and Uber (1998) and is referred to as the max-sum. The drawback of this approach is that the two most sensitive instrument locations could detect exactly the same set of events meaning that two sensors would be just as effective as one. Wu and Song (2012) also used a very similar approach however they accounted for the number of distinct events which are detected as well as adding a penalty term to ensure that the more robust sensor configuration is selected if the same number of events are detected for two configurations.

Related to the number of events which are detected is the detection probability. This is functionally the same, however it is normalised by the number of events which have been considered. Hagos et al. (2016) utilised the detection probability to select the sensor configuration and also performed further analysis on the chosen configurations to examine two other performance related metrics. The other two metrics were the false alarm rate and the average time taken to detect the leak/burst events. Rather than just analysing burst scenarios, as did Farley et al. (2008, 2010) and Wu and Song (2012), Hagos et al. (2016) generated normal data to determine how effectively a combination of pressure and flow meters could distinguish between the normal and burst data. The best configurations were those which provided the highest detection probability. The authors tried various different numbers of both pressure and flow sensors and determined the corresponding values for both the false positive rate and the detection probability. The authors noted that in future it would be worth simultaneously considering the detection probability and the false alarm rate to select the configurations. They also suggested that localisation could be formulated as part of this problem as well.

Puleo et al. (2018) used the identifiability index to measure the performance of each sensor configuration. The identifiability index is the ratio of two quantities derived from the sensitivity matrix. The normalised determinant of the parameters estimated at the middle of the parameter range is divided by the smallest root of the largest ratio of the eigenvalues of the Fisher Information Matrix (FIM). This is a measure of the sensitivity of the pressure at certain nodes to changes in the input parameters of the hydraulic model and is to be maximised to identify the best locations for placing sensors.

Raei et al. (2019) utilised the average time to detect all leaks/bursts to select the best pressure sensor configurations. The sensor configuration with the shortest average time to detect the leaks/bursts was taken as the best. Different leak/burst start times were used to ensure that the selected configurations are sensitive to events starting at any time of the day. For each leak/burst event a matrix is created containing the changes in pressure for all timesteps and at all nodes. The nodes are ranked according to the magnitude of the change in pressure across all timesteps. This process is repeated for all leaks/bursts so that the best sensor locations for all leaks/bursts can be determined.

Finally, Forconi et al. (2017) developed several risk-based objectives to evaluate sensor configurations. The first risk-based score was purely based upon the likelihood of detecting an event or not. This is equivalent to the previously described probability-based approaches or the max-sum approach outlined earlier in this section. Risk, however, is comprised of another element, which is the impact of an event which occurs. The risk is normally (in engineering practise) considered as the product of the likelihood and the impact of an event occurring. The second risk-based score proposed by Forconi et al. (2017) incorporated the impact by using a score derived from the sensitivity matrix. For each combination of node and leak/burst location the sensitivities were summed across all timesteps. The resulting values were then normalised by subtracting the minimum value in the matrix from all values and dividing by the range of values in the original matrix to force all values to be between 1 and 0. These values were then multiplied by the likelihoods used in the first score to give the overall risk. A further addition was made to account for vulnerable users where each network location was weighted according to the criticality of the demand at that point. This ensures that there is less tolerance of undetected leak/burst events in these areas. Independent of the risk score the max-sum was used to minimise the risk of leak/burst events by placing sensors to cover the area(s) of highest risk.

In terms of the constraints which were used to determine the optimal sensor locations there was variation in the approaches used. By far the most common approach is to use all pipes or some subset thereof as potential flow sensors and to use all junctions as potential pressure sensor locations. All of the studies in this section followed this convention. Junctions are hydraulic model components which represent the vertex between two links (pipes, pumps or valves) in a hydraulic model. Junctions can also have a demand, which represents customer demands, or an emitter coefficient which is used to simulate the release of water from the hydraulic model in order to mimic leak/burst events.

For modelling leak/burst events there was some variability between the techniques. As with the sensor locations all junctions were treated as potential leak/burst locations and then the same size, or range of sizes, of leak/burst event were modelled at each location. Farley et al. (2008, 2010), Huang et al. (2012) and Raei et al. (2018, 2019) used this approach. Forconi et al. (2017) and Puleo et al. (2018) modelled the leak/burst events at new junctions located at the midpoint in pipes rather than using the junctions which were in the hydraulic model. Wu and Song (2012) and Hagos et al. (2016) used Monte Carlo simulation to generate a set of leak/burst events with random size and location. All studies also assumed that only one event could occur at a time, except for Wu and Song (2012) who modelled two simultaneous events for each leak/burst scenario that was generated. However, it must be stated that this technique also uses model calibration and is aimed at historical background type leaks. Without this, it would not be possible to locate two events at the same time.

The most variability between the techniques occurs in the generation and evaluation of the sensor configurations so that the required level of performance against the defined objective(s) can be achieved. The simplest was the “brute force” or full enumeration approach which generates every possible combination of sensors and then evaluates each of them. This guarantees that the optimal solution will be achieved but required much more computational effort and leads to severe constraints which limit the scope of the techniques. Farley et al. (2008) used full enumeration but a limit of two sensors was enforced to keep the computational time reasonable. In order to expand the number of sensors (to allow 3 sensors to be placed) Farley et al. (2010) also employed a GA. GAs do not guarantee the optimal solution will be found but can quickly and efficiently explore the solution space to identify near optimal solutions if they are well designed. Raei et al. (2019) employed the NSGA-II GA for the same reason. This allowed two objectives to be simultaneously optimised, in this case minimising the number of sensors whilst maximising the performance. The benefit of such an approach is that the trade-off between the two objectives can be visualised and presented to decision makers so that more informed strategic decisions can be made. Raei et al. (2018) also used NSGA-II but in addition used k-means clustering to group together similar sensor locations prior to the determination of the optimal configuration to reduce the size of the search space in order to reduce the computational effort required. Wu and Song (2012) employed the Darwin Framework (Wu et al., 2011), in a similar way to determine the optimal sensor configurations. Hagos et al. (2016) used a general reduced gradient nonlinear solver (Mukai and Polak, 1976) to determine the best sensor configurations when the number of sensors was fixed. A slightly different approach was taken by Puleo et al. (2018), whereby the most sensitive nodes were first selected by averaging the sensitivities over all input parameters. From this selection of nodes, the best combination of nodes was determined using the identifiability index. It was not stated how the sensor combinations were selected from the subset determined by the sensitivity analysis. Huang et al. (2012) employed 3 different self-organising map structures to determine the optimal sensor configurations. Regardless of the input structure which was selected all leak/burst events were detected successfully. However, the number of sensors was not minimised directly and did not vary significantly.

2.4.2.3. Localisation

Localisation is the process, following detection, where a water company determines the approximate location of a leak/burst event. As with detection there is much variation in the literature in how the sensor placement techniques vary to achieve this. A case can be made that sensor placement, in reality, is a multi-objective optimisation problem. There are many individual objectives, some which are competing, which must be satisfied simultaneously. Therefore, selecting the objectives which will be used to evaluate the quality of any given sensor configuration is a critically important step in developing a sensor placement technique which reflects the reality of the leak/burst localisation sensor placement problem.

In relation to the first step in Figure 2-1, there are almost as many different objective formulations as there are sensor placement techniques. Sensor placement techniques even vary in relation to how many objectives are considered. The most common objective to consider is the number of sensors. This is logical as the cost associated with installing sensors in a water distribution system is closely related to the number of sensors which are installed. An assumption is generally used at this point which is that there is a linear relationship between the number of sensors which are installed and the cost of each sensor. In reality, there will be economies of scale which can be achieved for both the sensor and the installation cost but quantifying this, particularly in relation to the cost of installation is difficult, therefore it is ignored. When the number of sensors is used, generally it is given a fixed value and the optimisation in the other objective(s) is run for each number of sensors separately. This approach was used by Perez et al. (2009, 2011), Farley et al. (2013), Casillas et al. (2013, 2015), Sarrate et al. (2014), Soldevila et al. (2019), Steffelbauer et al. (2014), Steffelbauer and Fuchs Hanusch (2016) and Cuguero-Escofet et al. (2015). In reality, there is no reason why the number of sensors needs to be fixed and separate optimisation runs completed. In the cases where many objective functions were being simultaneously optimised then this would help to reduce the number of objective functions by one but for each of the studies listed above there is only one other objective being considered. This effectively reformulated the optimisation technique as a single-objective optimisation problem repeated several times rather than a multi-objective optimisation problem run only once. It is more efficient in terms of run time and user input to perform one multi-objective run as opposed to performing several single objective runs although the overall time taken for multiple runs may still only be in the order of hours (for typical size networks), in either case.

For each of the studies listed above, aside from the number of sensors to represent cost, the second objective is most often formulated as a measure of the localisation performance of the sensor configurations which are being evaluated. Between the sensor placement techniques this is where most of the variation between the techniques becomes apparent. For example, Perez et al. (2009, 2011) developed a measure of localisation performance which determined the ability of a sensor configuration to divide the leak/burst events into groups according to their leak signature. The leak signature, in this case, is a binary combination of the sensor responses in relation to an arbitrary threshold. Where the sensor response is higher than the threshold then a value of 1 is assigned and when it is lower, then a 0 is assigned. The localisation performance objective was calculated as size of the largest group of leak/burst event and was to be minimised to try and create groups which are as even in size as possible. Farley et al. (2013) used a similar objective based upon the size of the groups but considered the size of all groups in relation to the ideal size of group. As a binary leak signature is used there are 2^{n_s} possible groups, where n_s is the number of sensors being considered for placement. The ideal group size is the total number of nodes in the hydraulic model divided by the number of possible groups. Farley et al. (2013) minimised the total difference, for all groups produced by a given sensor configuration, between the group sizes and the ideal group size. Conceptually, this is very similar to Perez et al. (2009, 2011), however it considers all groups rather than just the largest group. The main advantage of both of these objective functions is that, assuming all leaks are detected and therefore correctly localised, the percentage of the DMA which must be searched, on average, can be determined easily.

Casillas et al. (2013) proposed an error index to measure how many events were not perfectly localised. In this case perfectly localised means that only the location of the leak/burst (and no others) was determined by the localisation technique. The error index was given a value of 0 when perfect localisation was achieved and a value of 1 when it was not. The average of the error indices for all considered leak/burst events was used as the objective for each considered sensor configuration. The authors also proposed another error index (Casillas et al. 2015) which determined the amount of overlap between regions on a hyperplane after the residual vectors for each leak/burst event was projected onto the hyperplane (in the leak signature space). The region of overlap is where leaks/bursts cannot be localised perfectly, because they can simultaneously belong to multiple groups with different leak signatures. The error index is to be localised such that the largest number of leak/burst locations belong to only one group, effectively minimising the overlap between the groups in the leak signature space. For both of these error indices calculating the actual localisation performance (as a measure of the size of the network) is more difficult because not all leaks/bursts are correctly localised, as they were for both Perez et al. (2009, 2011) and Farley et al. (2013). Steffelbauer et al. (2014) and Steffelbauer and Fuchs-Hanusch, (2016) proposed a modification to the error index proposed by Casillas et al. (2013)

which also accounted for process noise (caused by uncertainties in the demand) to penalise the performance objective. This has the effect of steering the sensor placement technique away from selecting sensor locations which suffer from high levels of process noise. Another modification proposed by Cuguero-Escofet et al. (2015) to the error index proposed by Casillas et al. (2013) was to account for model uncertainties by allowing a buffer distance around the actual leak/burst location. This meant that rather than the single leak/burst location being identified for the leak/burst to be considered correctly localised, that the identified leak/bust location just had to be in the region of the actual location. This is an important consideration for application of the sensor placement to real systems. Soldevila et al. (2019) presented an objective formulation for sensor placement based upon the error of a geostatistical technique (OK in this case) for determining the approximate location of a leak/burst event. As per previous studies, a fixed number of sensors was considered.

There are techniques which have been developed which handle the formulation of the objectives differently. Firstly, Sarrate et al. (2014) used the isolability index, which is similar to the error index used by Casillas et al. (2013). The objective formulation was inverted where the number of sensors was to be minimised and the required performance was set to a fixed value. The number of sensors was initially set to the number of valid sensor locations and a clustering approach was used to group the sensors according to the similarity of their responses to faults to reduce the starting number of sensors. The search aims to find the configuration with the minimum number of sensors which can achieve the required level of performance. This reflects the reality of the situation better than fixing the number of sensors and maximising the performance as generally, from an industry point of view, the required (or desired) level of performance will be known and should drive the sensor placement.

There are sensor placement techniques which simultaneously optimise multiple objectives. The advantage of using multiple objectives is that there is no need to specify directly the preference for either of the objectives before determining the trade-off curve (Pareto front) between the two objectives. Instead, the trade-off between the two can be calculated and presented to decision makers and then the entire trade-off curve can be used to select the preferred strategy (in this case the preferred number of sensors vs level of localisation performance). This was highlighted by Kapelan et al. (2005) in relation to hydraulic model calibration. This can also be achieved using a single run of the sensor placement technique, which is an additional benefit, although this may or may not be important depending on the time taken to complete each run for a given sensor placement technique. There are few techniques which simultaneously optimise two objectives in comparison to the number which optimise only one at any time.

Blesa et al. (2016) deviated from the majority of the sensor placement techniques in that they were optimising two performance measures for a fixed number of sensors. Both of the performance measures were related to the robustness of a sensor configuration which are based upon to the error index which was proposed by Casillas et al. (2013). The first measure was the same as that proposed in Casillas et al. (2013), which is the mean of the error indices across multiple leak/burst event sizes, and the second was the minimum of the individual error indices across multiple leak/burst event sizes. Effectively, both measures were looking to ensure that both the average performance and the worst performance were maximised, considering the multiple leak/burst event sizes. However, for each of these measures deriving a location accuracy (in terms of the percentage of the DMA which needs to be searched) is not possible directly. Again, this still has the problem with a requirement to specify the number of sensors prior to determining the sensor configuration.

Candelieri et al. (2014) proposed several bi-objective formulations of the sensor placement. One objective was the combined cost of flow and pressure instruments. The other objectives were either (i) the leakage index, which is the average size of the clustered leak/burst locations as a proportion of the size of the network, (ii) the quality of localisation, which describes the ability to cluster together the same leak/burst location across multiple leak/burst event sizes and (iii) the modified leakage index which is the product of the leakage index and the quality of localisation. This was the first example of a multi-objective formulation of the objectives for sensor placement which includes one objective for cost and one for performance.

In addition to the objective formulations which were described above it is necessary to define the constraints which are used to select the sensor locations and the leak/burst events which will be considered to determine the optimal sensor configurations. Of the two, determining the candidate sensor locations is generally the most straightforward. The most common approach taken in the literature is to consider the nodes (or some subset thereof) in the hydraulic model as potential sensor locations. Some techniques (Perez et al. (2009, 2011), Farley et al. (2013), Casillas et al. (2013, 2015), Soldevila et al. (2019), Steffelbauer et al. (2014), Steffelbauer and Fuchs-Hanusch (2016) and Cuguero-Escofet et al. (2015)) consider all junctions, that is all nodes not including reservoirs or tanks, in a hydraulic model as potential sensor locations. Alternatively, Sarrate et al. (2014) considered only the junctions which do not already have a demand (also referred to as dummy junctions) as candidate sensor locations. Blesa et al. (2016) took the opposite approach of considering only the junctions which do have demand as candidate sensor locations. The only study which considered pipes as potential candidate locations, because it was the only study included here which was designed to locate flow sensors alongside pressure sensors, was Candelieri et al. (2014). Candelieri et al. (2014) considered all junctions as candidate sensor locations.

To determine the range of leak sizes which are considered by a sensor placement technique several methods, ranging in complexity, have been used. The simplest is to use a single fixed size of leak/burst event at all valid leak/burst locations. The single size of event can either be added as a base demand (Perez et al. (2009, 2011), Cuguero-Escofet at al. (2015), Sarrate et al. (2014)) or is more commonly modelled using an emitter (Farley et al. (2013), Steffelbauer et al. (2014), Steffelbauer and Fuchs Hanusch (2016)). The difference between these approaches is that an emitter accounts for the nodal pressure in determining the outflow for a given leak/burst location whereas a base demand will be fully satisfied, even when there is not sufficient pressure to satisfy it. For this reason, an emitter will give a more realistic picture because the outflow will vary according to the location that the leak/burst event is being modelled. All of these studies can be considered limited in some way. This is because only one size of event has been considered and the optimal (or near optimal) sensor configurations which are derived from them can only be considered optimal with respect to those leak/burst event sizes. To deal with this problem some techniques include multiple leak/burst event sizes. Casillas et al. (2013, 2015), Blesa et al. (2016) and Candelieri et al. (2014) all varied the emitter coefficient for every leak/burst event location. The main problem with these approaches is that the range of event sizes must be determined manually by the user prior to running the sensor placement technique. This can require a detailed sensitivity exercise (or specific domain knowledge) for each DMA or WDS which is considered. If a large number of DMAs are being considered at any one time (which is a realistic scenario to consider) this would exacerbate the problem. When they are specified, there is no information about how the event sizes have been arrived at and what the main considerations are for selecting the leak/burst event sizes. For example, each event must be measurable (i.e. it causes a change at least equal to the magnitude of the sensor accuracy) by at least one sensor in order to be localised and this should be accounted for when the leak/burst event sizes are defined prior to running the sensor placement technique. This is not accounted for by any of the techniques and is seen as a major opportunity for further research.

The next two steps for sensor placement techniques, which are generating the sensor configurations and evaluating the performance of the configurations, are considered together. This is where most of the variation between the different sensor placement techniques can be found. Both steps together can be thought of as the process of converting the hydraulic data into the objectives, which have been discussed, and then systematically generating new sensor configurations to improve the performance against the objectives. The starting point of this process is a sensitivity matrix or several sensitivity matrices which are generated for all leak/burst event sizes at all leak/burst event locations. The changes in pressure for all candidate sensor locations are stored within the matrices for determining the optimal configuration. The first group of techniques to consider are those employing evolutionary algorithms, which is the most common approach. Generally, these techniques will use a localisation technique in combination with an evolutionary algorithm to maximise the localisation performance objective. Perez et al. (2009) started by normalising a single sensitivity matrix using the largest change in pressure for each leak/burst event (so that one node per leak/burst event has a value of 1 and the others have a value between 0 and 1) and the binarising (converting to a series of 1s and 0s) using an arbitrary threshold. The normalised, binarised sensitivity matrix is then searched using a single objective GA to identify the configuration which groups the leak/burst events most evenly according to the leak signature. This process is repeated for different numbers of sensors (between 2-11 in this case). The selection of the binarization threshold requires investigation and the authors trialled several different values which did not affect the performance significantly. However, the best threshold depends upon the number of sensors which are used which complicates the issue of selecting the best one. Perez et al. (2011) used the same fundamental approach whilst using different binarisation threshold values. The sensor placement technique was not applied to any real data using the optimal sensor configurations which needs further exploration to confirm the performance which was seen using the modelled data. Also, this technique compares the modelled data with the real data to localise real leaks/bursts which means that the model must be well calibrated.

Farley et al. (2013) utilised a similar approach which aimed for even sub-division of a DMA using the leak signatures for all modelled events. They too used a GA to search the binarised sensitivity matrix. The sensitivity matrix was populated with the X^2 error between the normal conditions and the leak condition instead of the absolute difference which was used by Perez et al. (2009, 2011). Farley et al. (2013) only considered either 2 or 3 additional instruments for a DMA producing either 4 or 8 sub-divisions. It is worth noting that for only 2 instruments a "brute force" search approach, where all possible sensor configurations are searched, was used. For 3 sensors a GA was used. In addition to accounting for the even sub-division a penalty was also applied to sensor configurations which produced large uncertainty zones. The uncertainty zone was applied either size of the binarisation threshold to account for the lack of uncertainty in the leak/burst simulations. The localisation for this technique is achieved by comparing the real data with itself and as such the hydraulic model does not need to be highly calibrated.

Casillas et al. (2013) proposed two different approaches for determining the optimal sensor configurations. Firstly, a semi-exhaustive approach, which systematically increased the number of sensors (adding 1 at a time) and then compared each new solution (with n_s sensors) with the previous best (with n_s-1 sensors) to determine if it performs better. If so, then the search is terminated. This can be repeated until as many sensors are placed as needed. The authors also used a GA to search for the optimal sensor configuration. In both cases the objective was calculated by considering pairs of residual vectors and sensitivity vectors which were created using different sized events. The projection of each residual vector against each sensitivity vector determines whether a leak is isolable (it can be successfully localised) and then the search techniques are looking to maximise the number of leaks which are localisable by minimising the error index. Robustness was considered by using time horizon analysis (considering multiple timesteps), use of a distance score (to allow the objective to be based on the quality of localisation rather than the binary error index) and by considering multiple different combinations of leak sizes to calculate the residuals and the sensitivities. The authors (Casillas et al. (2015)) also proposed a different formulation to remove the dependence upon the magnitude of leak/burst event by using vector projections onto a hyperplane under the assumption of linear approximations to the system of equations governing fluid flow. The sensor configurations, for multiple leak/burst event sizes, were then optimised for a given hyperplane which is selected for each number of sensors.

Sarrate et al. (2014) used a structural approach whereby the problem is transformed to a bipartite graph comprised of a set of equations, a set of unknown variables and a set of edges which describe the relationship between the model equations and the unknown variables. This graph can then be partitioned into smaller graphs, some of which are overdetermined, and these overdetermined subsets of the model equations can be analysed in terms of fault detectability and isolability. The sensor placement problem then becomes searching through the nodes in the graph, which relate to sensor locations, which gives the best isolability performance, for the desired number of sensors. The search was started at a given root node and then the graph is searched systematically to find new sensor locations which best the previous best identified isolability performance. The search was terminated if no improvement in performance can be found or the desired number of sensors cannot be installed. A clustering step is also used prior to the search to group sensor locations which are highly similar so that they are not duplicated in the selected configuration.

Blesa et al. (2016) utilised the sensor placement scheme proposed by Casillas et al. (2013), which is based upon the projection of a residual vector with each leak/burst event in the sensitivity matrix. However, instead of using a semi-exhaustive search or a GA/particle swarm optimisation (PSO - Kennedy and Eberhart, 1995) technique, Blesa et al. (2016) used a combination of clustering followed by an exhaustive search on the clustered candidate sensor locations. PSO is an evolutionary algorithm which mimics the behaviour of swarming animals to identify the best solutions to problems. A population of solutions is generated and moved around the solution space according to its best-known position as well as the best-known positions of other members of the population. The idea is that the population will converge towards to best solutions. A modified error index, proposed by Casillas et al. (2013), was used by Blesa et al. (2016) and was called the leak locatability index. This is calculated as the error index subtracted from the value of 1 and is to be maximised as opposed to minimised. Two statistics based upon the leak locatability index were used to determine the optimal sensor configurations using an exhaustive search.

Soldevila et al. (2019) employed a floating forward search (Pudil et al. 1994) to find the sensor configurations which minimised the error for an OK technique which is used to build up a map of pressure in all nodes in the DMA to determine the leak/burst location. The authors did not account for whether the leaks were correctly localised or not as there is a trade-off between the error of the OK technique and also the loss of information due to the limited number of sampling points being used. This is seen as an area for further improvement.

Steffelbauer et al. (2014) incorporated a differential evolutionary algorithm (Storn and Price, 1997) to solve the inverse problem in order to identify the leak/burst location based upon the pressure residuals calculated as the difference between a well calibrated hydraulic model and the data collected from sensors in a real WDS. Differential evolution was also used to determine the sensor locations to minimise the error index (Casillas et al. 2013). The error index was modified to account for near misses (i.e. where the actual location is very close to the determined location) by using different values based upon the proximity of the determined location to the actual location. The technique also incorporated process noise into the sensor placement technique by using a penalty in the projection between the residuals and the sensitivities vectors. However, it must be noted that the localisation technique (using differential evolution) is not the same as the sensor placement technique (using projection) which means that there is a discrepancy and that the sensor locations are not optimal with respect to the localisation technique being proposed. Steffelbauer and Fuchs-Hanusch (2016) used many of the same elements except that differential evolution for calculating the sensor configurations was replaced with a different GA (Nicklow et al. 2010). It was noted that there are several parameters which must be set according to user experience and that these could be determined using the GA at the same time as the sensor configurations. Cuguero-Escofet et al. (2015) again used a GA in combination with a modified error index calculated using the projection between the sensitivity vectors and a residual vector.

Candelieri et al. (2014) used an approach which was the combination of many techniques. The hydraulic simulations were conducted in a similar manner to Perez et al. (2009) except that the events were placed in each pipe rather than at nodes. In practise, it is unlikely that for a given event size this would drastically change the sensitivities measured at the nodes but the flow through the pipes would be affected which is why this approach is most suitable when flow sensors are considered. The range of event sizes selected did not consider the precision of the instruments being used and were specified in a range of allowable values at each location, using the emitter coefficient. A significant novelty in this paper is that the leak/burst events were grouped according to their similarity to reduce the number which are considered by the sensor placement technique. Spectral clustering (Von Luxborg, 2007) was used for this purpose. In this case the clustered leak/burst events are used to train a SVM to classify new leak/burst events which is how localisation is performed. Sensor placement is achieved by considering only the centroids of the clustered leak/burst events as the valid locations for either pressure or flow sensors. The number of sensors to be deployed is used to specify how many clusters the leak/burst scenarios are divided into. No tests have been completed using real data in conjunction with this technique. Given that the SVM is being trained using hydraulic model data the model will likely need to be well calibrated which adds to computational (or manual) effort associated with this technique and limits general applicability.

2.4.3. Hydraulic Model Calibration

2.4.3.1. Overview

Hydraulic model calibration is the process of updating or modifying a hydraulic model so that it more accurately reflects the system of which it is a model. This has been an active field of research for many years and many techniques have been developed for this purpose. Savic et al. (2009) reviewed techniques from the previous 30 years of hydraulic model calibration techniques and highlighted some key future areas for development of future techniques. A central point to note is that the requirement for calibration, and indeed the level of calibration, depends upon the purpose of the hydraulic model being used. Savic et al. (2009) highlighted that hydraulic models which are used for modelling water quality must be accurate, whereas a model required for making strategic decisions may not need to be to the same extent. Given that the level of calibration of a hydraulic model will also be related to the level of resources and personnel which are required to achieve this calibration it is clear why the use of advanced hydraulic model calibration techniques within water companies does not abound even though it has been a popular field of research. Sensor placement techniques for hydraulic model calibration, which aim to provide the best trade-off between the cost of sensors and the accuracy of the hydraulic model following calibration, are also common and have been formulated in a several ways. Generally, the parameters of the water quality model are adjusted so that the model outputs agree sufficiently with data collected from a real WDS. Kapelan et al. (2003) were the first to formulate the sensor placement problem for hydraulic model calibration as a multi-objective optimisation problem (Savic et al. 2009). In this case the number of sensors were minimised whilst the accuracy of the hydraulic model was maximised. Some other key points were also highlighted in relation to all sensor placement problems. Firstly, that an optimal solution with n sensors does not always contain the optimal solution with n-1 sensors. This has been assumed for a number of sensor placement techniques which use “greedy” type algorithms to reduce the amount of computational effort. Having said this, solutions which are near-optimal can often be achieved at a fraction of the computational effort. Also, by using multi-objective optimisation techniques the full range of performance for the objectives can be visualised and compared, allowing it to be presented to decision makers which can allow them to make more informed decisions.

Romano (2019) also highlighted the issue of hydraulic model calibration in relation to sensor placement for leak/burst detection and localisation. The key factor as to whether hydraulic model calibration was used was whether the hydraulic model was used to calculate the pressure residuals when the localisation step was performed on real leak/burst events or not. If the hydraulic model was used (for example by Casillas et al. (2013, 2015)) then hydraulic model calibration is required because the model error can easily outweigh the legitimate change in pressure due to a leak/burst event. This was the case for most of the techniques which were reviewed. Other techniques have been developed which do not use the model (such as Soldevila et al. (2019)) which means that the hydraulic model calibration becomes less important. There is still a need to determine an appropriate level of hydraulic model calibration and also compare how the effectiveness of sensor placement and leak/burst detection and localisation techniques are affected by the level of calibration which has been used. The current status of this issue is either that calibration is used or not used but the extent of the calibration (i.e. the average error between the model and the measured pressures corrected from the real system) is rarely quantified or explored. Using techniques which confine the model to the optimal sensor placement step but subsequently use a purely data-driven technique (comparing the current status with the past status using the real system only) for detection and localisation can alleviate the requirement for extensive hydraulic model calibration providing that models of sufficient quality are available. Finally, designing a sensor placement technique for both hydraulic model calibration and leak/burst detection and/or localisation, such as Wu and Song (2012) or Sophocleous et al. (2019a), can address this issue. The caveat to this is that more parameters need to be considered and a greater level of calibration accuracy is likely to be required in order to achieve this which increases the cost associated with this.

2.4.4. Water Quality/Contamination Event Detection and Localisation

Sensor placement techniques have been developed in order to detect and localise water quality or contamination events in water distribution systems. The purpose of these systems is to minimise the risk associated with the range of possible contamination events which can occur in order to protect customers (Hart and Murray, 2010). A review by Hart and Murray (2010) provided a comprehensive review of the range of techniques which have been developed to act as contamination event detection systems. The authors divided the techniques into three distinct categories; expert opinion, ranking methods and optimisation. Of these three categories the optimisation-based techniques offer the most promise for wide scale application to real WDS as they enable water companies to best strike the balance between two or more competing objectives. As with leak/burst detection and localisation one of the objectives is likely to be related to cost, as the same budgetary and resource constraints apply to contamination detection as for leak/burst events (UKWIR, 2013). The second objectives vary depending upon the type of events being detected but some examples are minimising public health impacts (number of individuals exposed or the number of individuals above a threshold of exposure), minimising the extent of exposure in the pipe network, minimising the detection time or minimising the number of events which were not detected (Hart and Murray, 2010). There are a number of parallels between the objectives used for contamination detection and for leak/burst detection, including some identical objectives used (e.g. number of sensors). On top of this, the optimisation-based approaches to sensor placement for contamination detection involve modelling a wide range of event types, sizes and locations and then placing sensors to best mitigate the risks across all event types.

There are major differences due to the differing nature of the events. Firstly, the types of sensors used for contamination detection are typically aimed at measuring one or more water quality parameters as opposed to measuring either pressure or flow. This is because contamination events can occur even if no change in the hydraulic conditions occur and pressure or flow changes can occur without contamination being present. On top of this, the risk of contamination is a spatial phenomenon which propagates through a water distribution system from a source or multiple sources.

Leaks/burst do not propagate in the same way but can cause widespread changes to the pressure and flow throughout a water distribution system. Several key issues not just related to the technical and computational challenge, but also related to the practical and feasibility of sensor placement in WDS were highlighted by Hart and Murray (2010) which are listed below.

- Real World Case Studies
    - In comparison with the number of case studies using modelled data there were few case studies which were applied to the real systems in order to validate the developed techniques. This leaves uncertainty as to whether techniques which performed well using modelled data could replicate this when utilised in an operational manner.

- Number of Sensors/ Multi-Objective Analysis
    - Typically, the optimisation-based sensor placement techniques only considered one number of sensors at a time and the trade-off between multiple objectives was not explicitly provided to decision makers. Using multi-objective techniques can provide the necessary information to allow better decisions to be made by personnel within water companies.

- Sensor Failures
    - It is an unavoidable fact that sensors will fail from time to time. Most of the sensor placement techniques for contamination detection considered a network of sensors which worked all the time. The effects of sensor failures need to be incorporated into the objectives which are considered to ensure robustness and to allow the risks of sensor failures to be accounted for. This can allow sensor configurations which are resilient to sensor failures (they do not suffer great reductions in performance when a sensor fails) to be identified. Section 4.4.2.3.7 provides further details on how this is handled by the leak/burst localisation technique.

- Dual-use Applications
    - It is of interest to water companies to explore how a sensor network can be used for multiple purposes to reduce the need for having entirely separate sensor networks. This can reduce the capital cost and ongoing maintenance costs of deploying two sensor networks.

- Heterogeneous Sensor Networks
    - It is a common assumption that the same type of sensor will be used to compose a sensor network. However, in the case of multiple different installations over time where there is an interest in extending the life of existing sensor infrastructure, this may not be the best approach. This is of more interest in the contamination detection sphere, where more parameters can be monitored, than for leak/burst detection and localisation.

Each of these areas represent a potential opportunity for improving the state-of-the-art in contamination detection and no technique has addressed all of the issues which were highlighted as being important. It is worth stating that aside from using heterogeneous sensor networks the other discussion points can also easily apply to sensor placement for leak/burst detection and localisation.

2.4.5. Water Quality Model Calibration

Of all of the sensor placement techniques aimed at WDS analysis the least common are those which aim to allow calibration of a water quality model. Generally, water quality models rely on a well calibrated hydraulic model (Savic et al. 2009) however there are additional parameters related to the interaction and reaction of certain chemical components in water which need to be determined. Examples of these include the bulk decay rate (Boccelli et al. 2003) and wall decay rate (Hallam et al. 2002) of chlorine travelling through the pipes in a WDS. The bulk decay rate can be taken as a constant value for an entire WDS or the value can vary according to the flow and water quality conditions in different parts of the system. The wall decay rate is a pipe-specific parameter which needs to be determined in order to accurately model the decay of chlorine throughout an entire system. This is one of the more common examples but the same principle (in theory) applies to modelling the accumulation or decay of any water quality parameter.

Given that the sensors which are placed for calibrating water quality models are likely to be monitoring water quality parameters (such as chlorine, turbidity, iron, or dissolved oxygen) it is not likely that the same techniques which are used to place these sensors will be relevant to placing either pressure or flow sensors. In addition to this, the locations which are selected for calibrating models (both hydraulic and water quality) will not be the same as those for detecting and localising events, even when the same types of sensors are used.

2.4.6. Summary of Sensor Placement Literature Review

The main conclusions from the sensor placement literature review are listed below:

- The performance of sensor placement techniques are coupled with any analytical techniques being used to analyse the data collected from the sensors that are being placed. Therefore, to determine the performance of any sensor configuration it is crucial that the analytical technique is also considered. For example, if sensors are being placed for leak/burst localisation then the localisation technique which will be used must also be considered during the selection of the sensor configurations to ensure that they perform optimally with respect to the localisation technique. This is a key opportunity for improvement in the current sensor placement literature.

- The sensor placement problem is inherently multi-objective. Generally, in an effort to reduce the complexity of the computational task, sensor placement techniques for multiple different purposes fix the objective related to cost and repeat multiple runs with different cost values. It is quicker and more convenient to consider multiple values for the cost and optimise this during a single sensor placement run. On top of this it determines the trade-off between the two (or more) objectives at once whilst alleviating the need to specify the constraints prior to running the sensor placement technique. This means that decision makers can consider more information to make better decisions.

- Following from this several different types of techniques for determining the optimal sensor configurations can be found. The problem is that the majority follow a “greedy” strategy, assuming that the adding a single sensor to an optimal sensor configuration with n sensors will give a new optimal configuration with n+1 sensors. In reality, “greedy” strategies can overlook optimal solutions although they are typically computationally efficient as fewer solutions need to be evaluated. In cases where “greedy” strategies are used, relatively few studies actually explore the effect that this has on the chosen solutions by comparing them with the true optimal solutions. Therefore, these techniques are not necessarily considered proven and further work is required in order to demonstrate them fully by comparing them with, for example, GAs.

- Some key practicalities related to the localisation of leaks/bursts are not accounted for currently by most sensor placement techniques. For example, the size of the search area (rather than just whether an event was correctly localised) is a key factor which impacts upon the quality of a sensor configuration.

- Selecting the size of leak/burst events to model for determining the optimal sensor configurations is often left down to the judgement of the end-user. Either a single size of leak/burst events is used or a very narrow range of event sizes are specified. The sizes of events should be targeted towards the type of events which cannot be localised by other means which are already being utilised by water companies to best target the available resources at the most event sizes which have been neglected to date.

- Related to the previous point the accuracy of the instruments should also be accounted for in the sensor placement technique. Nearly all of the techniques which were reviewed did not consider whether the events which are modelled could actually (in the real world) be measured with certainty, given the inherent variability which is ubiquitous in WDS. Similarly, many leak/burst events are likely to cause exactly the same effects when the accuracy of the instruments is accounted for. This means that there is much duplication in the events considered by the sensor placement techniques, increasing the computational time and biasing the sensor locations towards areas in the model with a large number of potential locations. Given that these locations do not match the potential locations in a real WDS this is problematic and needs to be addressed.

- Only one such review aimed to compare the sensor placement techniques with each other. However, there is a lack of quantitative assessment of the techniques against each other using benchmark hydraulic models and real data. The case of using real data would be difficult as it would potentially require a large number of sensors to be installed in a single DMA but there is no reason why at the current time the model performance cannot be used to compare the techniques. The major problem is that different objectives are used and some of the techniques would need to be modified so that a consistent metric can be used to compare the techniques on a like for like basis. The also highlights the need for a more consistent framework for sensor placement as there is for other issues in WDS analysis.

- Although many techniques use algorithms to automate the large amounts of data analysis required for sensor placement, there are still (in most cases) multiple parameters which need to be selected prior to being able to run the sensor placement technique.

- More realistic functions could be used for both the cost and performance objectives. Typically, the cost is assumed to only depend upon the number of sensors and the performance is related to the number of events correctly localised or detected. However, the quality of localisation (i.e. the accuracy) is very important and will be the most critical factor in determining how good a sensor configuration is.

Table 2-3 *Comparison of sensor placement objective functions for leak/burst localisation*

Reference	**Purpose**	**Objective Function 1**	**Objective Function 2**	**Optimisation**
Perez et al. (2009)	Detection/Localisation	N/A – Fixed number of sensors chosen (repeated)	Largest group containing events with all signatures (all events detected)	GA
Farley et al. (2013)	Localisation	N/A – Fixed number of sensors chosen (repeated)	Evenness of group sizes (total deviation from ideal size)	GA
Perez et al. (2011)	Localisation	N/A – Fixed number of sensors chosen (repeated) – flatness of cost curve	Largest group containing events with all signatures (all events detected)	GA
Farley et al. (2010)	Detection	N/A – Fixed number of sensors chosen (repeated) – 1 sensor/time	Max-Sum (Bush and Uber, 1998)	GA
Farley et al. (2008)	As above	As above	As above	Brute Force (Full enumeration)
Casillas et al. (2015)	Localisation	N/A – Fixed number of sensors chosen (repeated)	Error index (which is the overlap between the leak signatures where they are not isolable)	GA/PSO/Semi-Exhaustive
Casillas et al. (2013)	Localisation	N/A – fixed number of sensors chosen (repeated)	Error index (% of events not perfectly localised)	GA/PSO/Semi-Exhaustive
Sarrate et al. (2014)	Localisation	Number of Sensors	Above threshold for fault isolability index (fixed)	Branch & bound search with clustering
Soldevila et al. (2019)	Localisation	N/A – fixed number of sensors repeated	Sum of squared errors for kriging model	Sequential forward search
Blesa et al. (2016)	Localisation	Mean leak locatability index	Worst leak locatability index	Clustering and exhaustive search thereafter
Steffelbauer et al. (2014)	Localisation	N/A - fixed number of sensors (repeated)	Error index with penalty for high output noise	Differential Evolution
Steffelbauer and Fuchs-Hanusch (2016)	Localisation	N/A fixed number of sensors and fixed penalty (ω) – (repeated)	Error index with penalty for high output noise	GA
Cuguero-Escofet et al. (2015)	Localisation	N/A – fixed number of sensors (repeated)	Weighted cross-correlation between two leak/burst locations penalising high correlation for leaks which are far apart	GA
Candelieri et al. (2014)	Localisation	Cost (combined for pressure and flow sensors)	Localisation performance (either leakage index or quality of localisation index)	Clustering (non-optimised)

Table 2-4 *Comparison of sensor placement constraints and parameters for leak/burst localisation*

Reference	Purpose	Leak/Burst Event Scenarios	Valid Sensor Locations/Numbers
Perez et al. (2009)	Detection/Localisation	All junctions/constant demand/6l/s	All junctions/2-11 sensors
Farley et al. (2013)	Localisation	All junctions/emitter/not stated	All junctions/2 sensors
Perez et al. (2011)	Localisation	All junctions/constant demand/1l/s (3% night-time)	All junctions/2-11 sensors
Casillas et al. (2015)	Localisation	All junctions/emitter coefficient (varying)	All junctions/2-5 sensors
Casillas et al. (2013)	Localisation	All junctions/emitter coefficient (varying)	All junctions/2-10 sensors
Sarrate et al. (2014)	Localisation	All junctions without demand/6.3ls average (demand)	All junctions/8 sensors
Soldevila et al. (2019)	Localisation	All junctions/25-75 l/s (Hanoi)	All junctions/5 sensors
Blesa et al. (2016)	Localisation	All demand junctions/0.5-0.95l/s (Hanoi)/5 clusters for real DMA	All non-demand junctions /2-5 sensors
Steffelbauer et al. (2014)	Localisation	All junctions/emitter coefficient = 0.5 (sensitivities), 0.6 (residuals) – synthetic/ emitter coefficient = 1 (sensitivities), 1 (residuals) - real	All junctions/2 sensors (synthetic)/5 sensors (real)
Steffelbauer and Fuchs-Hanusch (2016)	Localisation	All junctions/emitter coefficient = 0.5	All junctions/2-5 sensors
Cuguero-Escofet et al. (2015)	Localisation	All junctions/6 l/s	All junctions/2-4 sensors
Candelieri et al. (2014)	Localisation	All pipes (0.15-0.1575) using emitter coefficient	Centroid of clusters (pipes/junctions)

Table 2-5 *Comparison of sensor placement objective functions for leak/burst detection*

Reference	Purpose	Objective Function 1	Objective Function 2	Optimisation
Farley et al. (2010)	Detection	N/A – Fixed number of sensors chosen (repeated) – 1 sensor/time	Max-Sum (Bush and Uber, 1998)	GA (up to 3 sensors)
Farley et al. (2008)	As above	N/A – Fixed number of sensors chosen (repeated) – 1 sensor/time	Sum of binarized X^2 error metric (adding 1 to previous best)	Brute Force (2 sensors only)
Wu and Song (2012)	Detection	Fixed number of sensors	Maximise the sum of distinct leak/burst events which can be detected + total number of event detections	Darwin optimisation framework (Bentley)
Hagos et al. (2016)	Detection	Fixed number of sensors	Maximise detection probability	General reduced gradient (Abadie, 1970)
Huang et al. (2012)	Detection	Not directly specified	Similarity between normalised response to all leak/burst events (used to cluster)	Self-Organising Map
Puleo et al. (2018)	Detection	N/A – Fixed number of sensors chosen (repeated)	Identifiability index (using selected candidate nodes ranked using Fisher Information Matrix)	Monte Carlo Simulations
Raei et al. (2018)	Detection	Minimise number of flow then number of pressure once flow fixed	"Value of identification" based upon allocating the leak/burst locations in the correct zone for chosen pressure and flow sensors	k-means clustering and NSGA-II
Raei et al. (2019)	Detection	Number of sensors (minimise)	Time to detection (minimise)	NSGA-II with different measurement errors for pressure measurements
Forconi et al. (2017)	Detection	Fixed number of sensors	Max-sum of 3 risk-based scores; (i)probability of detection, (ii) probability of non-detection x impact of non-detection, (iii) weighted towards vulnerable users	Not stated (assuming taking the n sensors with largest risk score)

***Table 2-6** Comparison of sensor placement constraints and parameters for leak/burst detection*

Reference	**Purpose**	**Leak/Burst Event Scenarios**	**Valid Sensor Locations/Numbers**
Farley et al. (2010)	Detection	All junctions/4 different leak diameters used to determine emitter coefficient (6, 10, 17 and 20mm)	All junctions
Farley et al. (2008)	As above	All junctions/5 different leak diameters used to determine emitter coefficient (2-10mm)	All junctions/max 2 sensors
Wu and Song (2012)	Detection	2 leaks simultaneously applied at random junctions using between 0.315-0.328 l/s at each location. 10000 scenarios generated (MCS)	5, 10, 15, 20, 25, 28 sensors at junctions
Hagos et al. (2016)	Detection	All junctions/100 burst events with random location and size (1-50 emitter coefficient) (MCS)	All junctions (1-125 pressure meters), all pipes (up to 90 flow sensors)
Huang et al. (2012)	Detection	10% of base demand for all junctions	All junctions
Puleo et al. (2018)	Detection	New junctions placed at mid-point of each pipe/Sizes not stated	Not stated (assuming all original junctions)
Raei et al. (2018)	Detection	All junctions (5-20gpm)	Not stated (assuming all pipes and junctions)
Raei et al. (2019)	Detection	All junctions 0.2-0.5l/s	All junctions
Forconi et al. (2017)	Detection	All pipes (centre point) 20% of inflow	Flow at all pipes, pressure at all junctions

Chapter 3 - Aims and Objectives

3.1. Aim

The aim of this thesis was to develop, test and validate a novel, integrated optimal sensor placement and leak/burst localisation framework which allows the approximate location of newly occurring leaks/bursts within a DMA. The first stage of running the integrated optimal sensor placement and leak/burst localisation framework determines the optimal configuration(s) of pressure sensors (in terms of number and location of sensors) within a DMA. The pressure sensor configurations capture pressure deviations caused by new leak/burst events to achieve the best level of localisation accuracy when used in conjunction with the leak/burst localisation technique which was developed as part of the framework. The optimal pressure sensor configurations with varying numbers of sensors, and the corresponding level of leak/burst localisation performance achieved for each, can be presented to water company personnel so that the preferred sensor configuration can be selected. Crucially, and in contrast to the current optimal sensor placement techniques identified in the literature, the optimal sensor configuration(s) are determined with respect to the leak/burst localisation technique developed for the integrated framework, to ensure that the configurations are optimal with respect to the leak/burst localisation technique. This is a significant advance on sensor placement techniques which only use even spacing or entropy (in the sense of information theory) to determine the optimal sensor configurations. Further to this, the optimal sensor configurations are determined with respect to multiple leak/burst event sizes which are determined automatically for each DMA. The second stage in running the integrated framework fuses multiple pressure signals captured by the sensors deployed in a DMA to determine the approximate location of the leak/burst using spatial/geostatistical techniques.

3.2. Objectives

The aim was broken down into 8 key objectives which are listed below. The objectives were highly interdependent and, as such, were achieved iteratively.

1. Select a set of performance metrics which are used by the sensor placement technique and leak/burst localisation technique to evaluate their performance during development, testing and validation.

2. Identify suitable constraints, and their values, for use by the sensor placement technique and the leak/burst localisation technique during development, testing and validation.

3. Understand and identify key factors which determine how the size and location of leak/burst events affect the variation in pressure within water distribution systems (WDS) using a combination of offline hydraulic modelling (with modifications made to ensure accurate modelling of leak/burst events) and engineered events (leaks/bursts simulated by opening fire hydrants).

4. Develop a technique which automatically determines a range of leak/burst event sizes, for each possible leak/burst event location. The minimum sizes shall ensure that the changes due to each leak/burst event are large enough to be measured by at least one sensor location. The maximum leak/burst event sizes will be limited by specifying a maximum increase in flow to ensure that the events are not large enough to be considered as reported events.

5. Develop and test a leak/burst localisation technique, which can quantify the magnitude of deviation (as opposed to just identifying that a significant change has occurred) in a pressure signal from a number of deployed sensors due to the presence of a new leak or burst by comparing the pressure signals collected prior to the initiation of the new leak/burst with those collected during the presence of a new leak/burst.

6. Develop novel, spatially-constrained (i.e. using an appropriate distance measure to replace Euclidean distance) interpolation/geostatistical technique(s) which combine the outputs of the technique developed in objectives 4 and 5 for multiple pressure sensors to approximate the pressure deviation for every location in a DMA. The search area is then inferred from the determined pressure deviations.

7. Develop a novel optimal sensor placement technique which, utilising the techniques developed in 4, 5 and 6 in combination with an optimisation technique, determines the optimal configuration(s) of additional pressure instrumentation in a DMA (such that the number of instruments is minimised whilst also minimising the size of the corresponding search area). This will allow selection of an optimal sensor configuration from the Pareto front which can achieve a desired level of localisation accuracy.

8. Field test/validate the novel, integrated optimal sensor placement and leak/burst localisation methodology developed in 4, 5, 6 and 7 using engineered events in a number of real DMAs. Pressure sensors will be deployed in the selected validation DMAs to determine how well the different sensor configurations are able to localise the engineered events.

Chapter 4 - Framework of Methods

4.1. Introduction

In this chapter a novel, integrated framework of methods to achieve the dual aims of optimal sensor placement and leak/burst localisation is presented. The framework of methods makes uses of a number of techniques, some of which have been newly developed for the framework to achieve the aims and objectives which were provided in Chapter 3. For each of the techniques which have been developed or adopted within the framework of methods the justification behind their inclusion will be described in addition to providing technical details related to their implementation. Due to the highly interconnected nature of the techniques within the framework of methods where there is interaction between two parts of the framework this will be described as well.

4.2. Rationale Behind the Framework of Methods

The framework of methods has been developed for the purpose of localising leak/burst events in a DMA. The two key steps, in the order which they are performed by the framework of methods are, firstly, to determine the optimal (or near optimal) configuration of pressure sensors by analysing a hydraulic model of a DMA and. subsequently. to analyse the pressure data from the optimal sensor configuration to determine the approximate location of a new leak/burst which has been detected in a WDS. A key point, which is novel to this framework of methods, is that the same analytical techniques are used during both the sensor placement step and in the leak/burst localisation step which ensures that the sensor locations determined by the sensor placement technique are optimal with respect to the leak/burst localisation technique which has been developed. The pressure sensor locations, determined using the framework of methods, are used for localisation only and the detection of leak/burst events is outside of the scope of the framework of methods. Once an event has been detected then the date of the event is passed to the leak/burst localisation technique so that the approximate location of the event can be determined.

Both the sensor placement and leak/burst localisation techniques rely upon the fact that the characteristics of a new leak/burst event will lead to a specific set of effects on the hydraulics of the DMA which can be measured and then exploited to infer the approximate location of the leak/burst event. There will be differences between the effects depending upon whether the event is modelled or is in a real WDS because a hydraulic model cannot perfectly match the real WDS. Pressure has been selected as the hydraulic variable to measure due to the much lower cost and ease of installation of pressure instruments relative to flow instruments. Most pressure instruments can be installed using a pressure tapping which means that access to only a portion of the pipe is required as opposed to flow instruments which require total excavation and isolation of a portion of pipe. A section of the pipe must be removed and the flow instrument inserted in place of the cut-out piece. Aside from the higher time cost of performing this work after the installation of the flow sensor, the isolated section of the pipe will need to be flushed and then disinfected to prevent contamination. In the case of pressure sensors, the cost of the instrumentation makes up a small proportion of the total installation cost. Combining this with the measurement uncertainty of deploying a single instrument at each location means that deploying multiple pressure sensors at each selected sensor location would reduce the measurement uncertainty and increase the resilience to failures. However, this would significantly increase the number of instruments required and is typically not desirable for a water company, particularly for a non-critical application like leak/burst event detection and localisation.

The current level of WDS monitoring makes detection (i.e. identifying the DMA(s) containing leak/burst events) possible but localisation (or narrowing down the area within a DMA containing the leak/burst event) requires a higher density of instrumentation than are currently deployed in most WDSs. Additional pressure instruments, and the data collected from them, are central to the developed framework of methods and it is crucial that the optimal (or near optimal) combination of locations for the pressure instruments are selected. This allows the minimum number of sensors to be used to be used by the framework of methods, to minimise the associated cost of the instruments for a water company whilst simultaneously maximising the performance of the deployed sensor configurations. This is important due to the tightening budgetary constraints which are imposed on water companies which often means that only expenditure with the highest ratio of benefit to cost can be justified, particularly for new technologies. The framework of methods makes use of a novel sensor placement technique to determine the optimal configuration of pressure sensors for a DMA so that the best leak/burst localisation performance can be achieved.

Once the best configuration of pressure sensors has been determined, using leak/burst simulations in a hydraulic model, and the pressure sensors have been deployed then the novel leak/burst localisation technique can be run when a new leak/burst is detected by the event recognition system used by United Utilities. The framework of methods can be used with any leak/burst event detection technique including the traditional approaches which are currently used by many water companies. The leak/burst localisation technique which has been developed can estimate the area (a set of DMA pipes) containing the new leak/burst event by comparing the night-time pressure signals after the leak/burst event has occurred with those from before the commencement of the leak/burst event. The magnitude of the change in pressure, measured by each deployed pressure sensor, is then used as the basis for leak/burst localisation. This means that there can be, at most, 24 hours between the commencement of the leak/burst event and the approximate location being calculated. Using the night-time pressure is important because it maximises the sensitivity of the pressure to the size of the leak/burst event and means that smaller events will produce larger relative changes in pressure which can be used to localise the leak/burst event. The framework of methods presented here is targeted towards events that would not be reported by customers which are typically referred to as unreported leaks/bursts. These events can easily be masked by the higher flows which are typically seen during the day but this problem is improved by using the night-time period to localise the leak/burst events.

Another important consideration of using the novel leak/burst localisation technique, which is a data-driven technique, is that pre-existing leak/burst events cannot be localised. This is because the changes in pressure will occurred prior to the deployment of the pressure sensors and therefore would not have been captured. However, in the overwhelming number of cases these events will be small and therefore would not significantly impact the hydraulics of a DMA, meaning that the changes in pressure would not have been large enough to be measured using pressure sensors. The remainder of this chapter shall outline the framework of methods, including the novel sensor placement technique and leak/burst localisation technique, which were developed to localise new leaks/bursts occurring in WDSs.

4.3. Development of the Framework of Methods

The goal for leak/burst localisation techniques is to use them in real WDSs. The data collected from real WDSs is often “noisy” when compared to data calculated using hydraulic models, which presents a challenge. Many leak/burst localisation techniques are developed for use with hydraulic model data only, in other words they only ever use simulated data. This can present problems when techniques which are developed using hydraulic model data are applied to data from real WDSs. To help overcome this issue, data collected from a real WDS was used during the development of the framework of methods presented in this thesis. The advantages of this are that the distinctive features which are common in real data are also considered during the development of the framework of methods rather than only at the end when testing is conducted. In order to achieve this engineered event data was collected at the outset of the project. Engineered events are synthetic leak/burst events which are created by opening fire hydrants or washouts to allow the controlled release of a quantity of water for a set period of time. The advantages of engineered events are that good control over the leak/burst characteristics can be achieved and that the location and timing of the engineered event can be known precisely. The flow rate of the engineered events can be controlled by using an inline flow meter in the standpipe which is used to direct the released water to a nearby drain. The hydrant or washout can be adjusted until a desired flow rate is achieved. Knowing the leak/burst characteristics makes evaluating the performance of the framework of methods much easier as it important to know the exact location of a leak/burst event in order to determine if it was localised correctly.

There is a tight coupling between the leak/burst localisation technique and the sensor placement technique. Therefore, whenever an update or change was proposed to one of the techniques then the same change was also made to the other. This necessitated an iterative approach whereby once changes were made and evaluated for one half of the framework of methods then they were made to the other half and the performance was evaluated once again. This prototyping process was repeated until satisfactory performance was achieved for all parts of the framework of methods.

4.4. Framework of Methods

The framework of methods presented in this thesis is outlined in the coming sections. Figure 4-1 depicts the final framework of methods, which was arrived at after several rounds of iterative development, in its entirety.

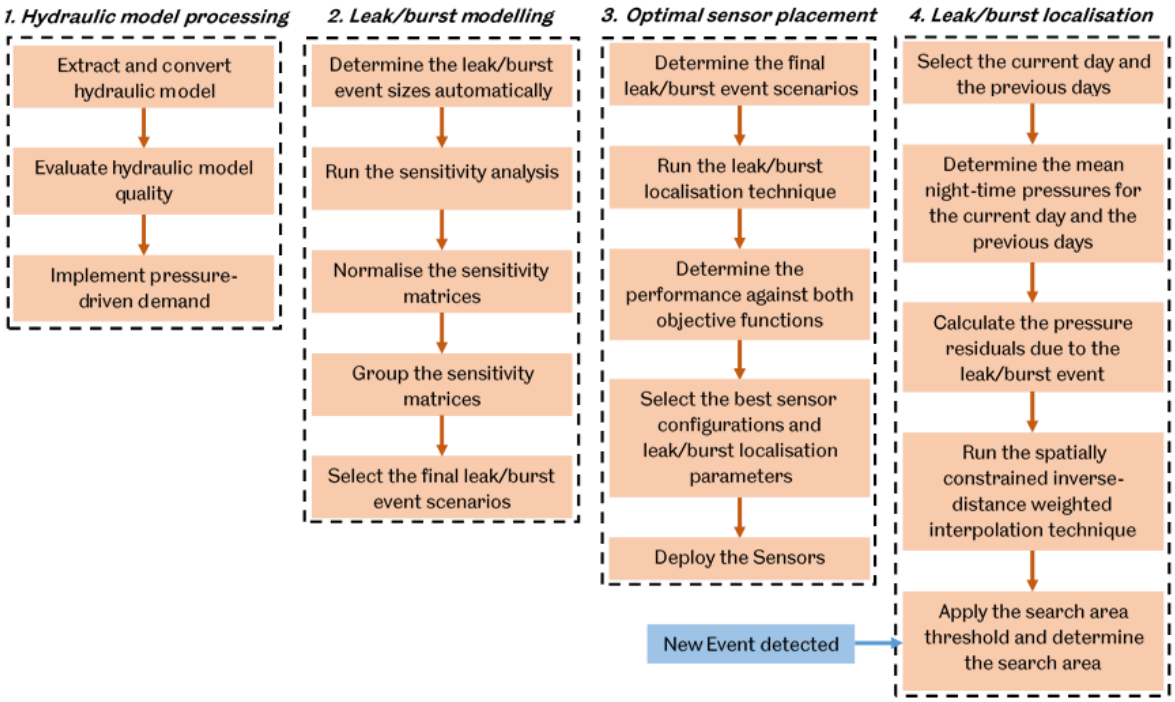


Figure 4-1 *Schematic of the framework of methods*

This includes all of the techniques which are used, the order in which they are performed and how they are connected to each other. The leak/burst localisation technique is used by the sensor placement technique therefore it is described first.

4.4.1. Leak/Burst Localisation Technique

4.4.1.1. Overview

The novel leak/burst localisation technique outlined here makes use of additional pressure instruments and a novel spatially constrained version of IDW (referred to hereafter as SC-IDW) interpolation technique (Zimmerman et al., 1999) to determine the location of new leaks/bursts occurring in a DMA. The leak/burst localisation technique analyses the data collected from the additional pressure instruments in a DMA over a number of days to determine how much change in pressure has occurred at each sensor location due to the new leak/burst. The novel, SC-IDW interpolation then estimates the changes in pressure at all locations throughout the DMA to identify the area with the highest change in pressure, which is considered as the most likely area in the DMA containing the new leak/burst event. Field teams can use this information to focus the search for the leak/burst event so that it can be precisely located and repaired. The key steps in performing the leak/burst localisation technique are outlined in the remainder of this section. Figure 4-2 outlines the comparison between leak/burst detection and leak/burst localisation.

In Figure 4-2 the two green instruments represent the regulatory flow and pressure instruments which are used to monitor every United Utilities DMA. A measurement is collected every 15 minutes and are analysed by the event recognition system which has been developed and implemented by United Utilities in recent years. The light green search area shows the area that would need to be searched once the leak/burst event has been detected using the event recognition system. The approach used by the framework of methods presented in this chapter is to deploy several pressure instruments throughout the DMA and collect a reading every minute. This allows the search area to be focussed around the area containing the leak/burst event as is shown by the purple area enveloping the two sensors in the right-hand side of the DMA.

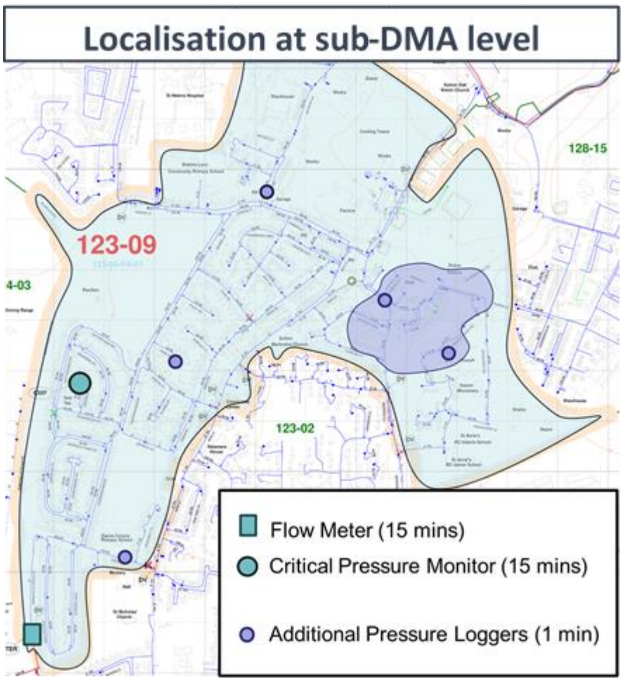


Figure 4-2 *Comparison between detection (using a single flow and pressure instrument) and localisation (using multiple pressure instruments) for a typical DMA*

4.4.1.2. Determining the Pressure Residuals

The main steps in determining the changes in pressure due to a new leak/burst event are shown in Figure 4-1. Pressure residuals denote the difference between the estimated value of a variable and the observed value. They are commonly used during model-based leak/burst localisation techniques to denote the difference between the model outputs and the measured pressure collected from the WDS and are used to infer the location of a leak/burst. This section will describe a data-driven approach to determining the pressure residuals for a new leak/burst event. The majority of leak/burst localisation techniques compare the pressures in a model with those collected from a real WDS to determine the changes in pressure which have occurred. The novelty in this technique is that it relies on comparing the historic pressures at the various sensor locations throughout the DMA with the pressures collected shortly after a leak/burst event has started.

The key to the approach is that data from hydraulic models are not used to localise real leak/burst events. The first step of this approach is to select the analysis window which will be used to determine both the normal pressures (with no leak/burst present) and the leak/burst pressures. Throughout the day the customer demands in a WDS vary and this causes variations in the pressure as a result. The flows (and hence pressures) typically follow a daily cycle which is referred to as the diurnal flow pattern. This pattern is regular and repeats, with some variation, every day. A pressure and flow plot showing several days of pressure and flow data for a DMA is shown in Figure 4-3.

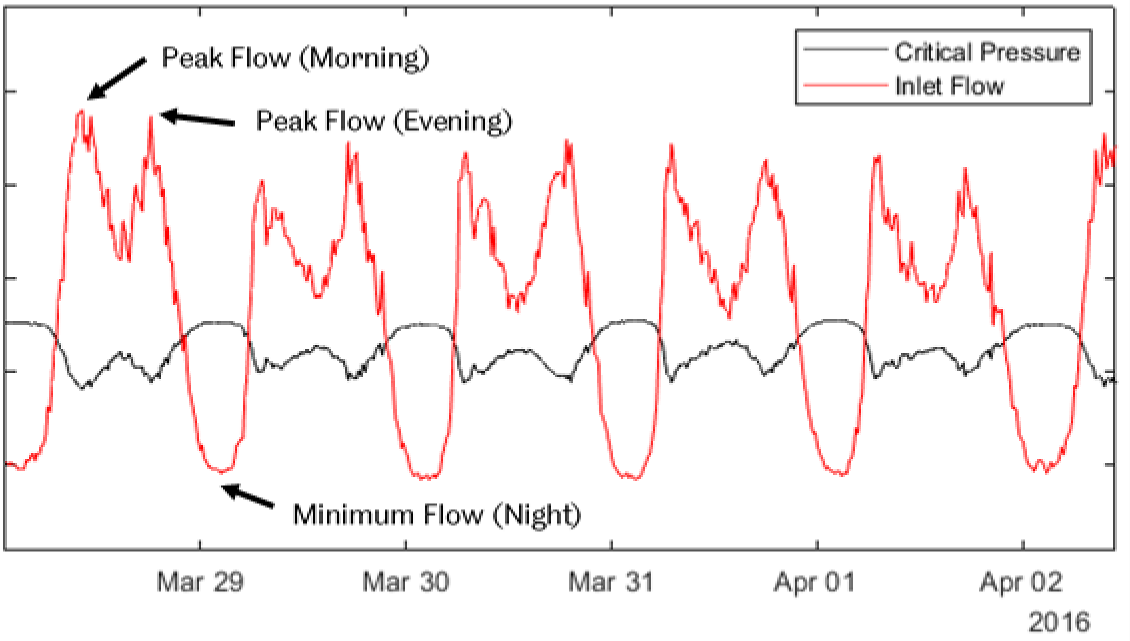


Figure 4-3 *A typical diurnal flow and pressure pattern in a gravity fed DMA*

The highest flows and hence minimum pressures occur in the morning and the evening each day and the minimum flow occurs during the night when most customers are asleep. The regularity of this repeating cycle is exploited to determine the pressure residuals. This is achieved by taking the average pressure, over a number of days prior to the leak/burst event being detected, to determine the normal pressure for each sensor location. The average pressure is calculated during the analysis window between 03:00 and 04:00, to maximise the sensitivity to leak/burst events. As can be seen from Figure 4-3 the pressure during the night hours generally varies much less than the pressures during the daytime. This is because the customer demand tends to vary less during the night. The leak/burst pressures are calculated during the same analysis window (between 03:00 and 04:00) on the day that the leak/burst event occurred. The pressure residuals are then calculated, for each sensor location, as the difference between the normal pressures and the leak/burst event pressures.

An important consideration for determining the pressure residuals is the number of days used to determine the normal pressures. Considering too many previous days could lead to the inclusion of operations which are normal but lead to significant changes in pressure. An example of this could be a rezone, where the boundary valves in the WDS are opened and closed to change the route of flow through the WDS. It is desirable to remove operations such as these as the changes in pressure could easily mask the presence of a new leak/burst event. Conversely, using too few previous days will increase the influence of the inherent variation of the customer demands on the average of the previous days and will also increase the influence of anomalous days, where the demand is much higher or lower for legitimate reasons. Previous data-driven approaches (see section 2.3.6), which rely on the comparing the past behaviour of the WDS with the present behaviour, have studied the effect of the number of previous days. The performance of the novel leak/burst localisation technique, presented in this thesis, is shown in Figure 4-4 for various numbers of previous days.

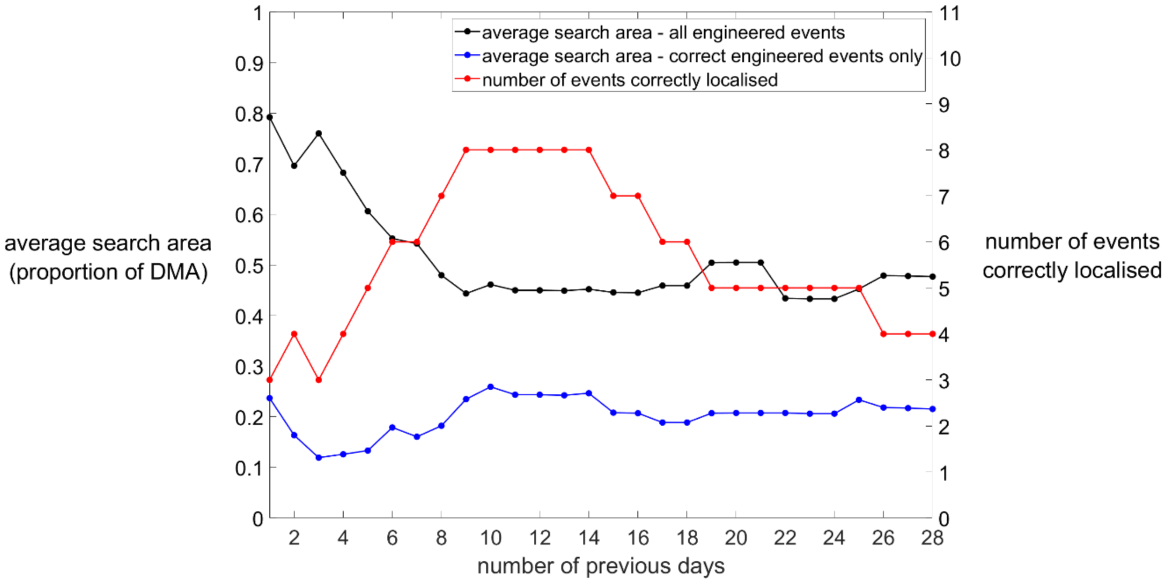


Figure 4-4 *Performance of the novel leak/burst localisation technique for various numbers of previous days*

Typically, using one or two weeks of data to form the previous day data set were used as the best trade-off. Following a sensitivity analysis of the performance of the leak/burst localisation technique for several engineered events the best localisation performance was found when using between 10 and 14 previous days for DMA 123-09. The best number of previous days to use depends upon the variability in pressure for a given DMA and must be determined for each DMA considered. The highest number of correctly localised leak/burst events occurred when between 10 and 14 previous days were considered. The effect on the average size of the search area, considering all of the engineered events, was that when more than approximately 10 previous days were used the average size of the search area was not significantly affected but the number of leak/burst events correctly localised was significantly worse (by between 9% and 45%).

It is desirable to remove any days from the previous day set on which leak/burst events were known to have occurred. To do this every time a leak/burst event is detected in a DMA the day on which it occurred is recorded and stored. When the previous day data set is assembled any event dates which were stored in the list of previous event days are removed. To compensate for the lost day(s) an earlier day(s) is added so that the chosen number of previous days is maintained. The final step in assembling the previous day data set is to separate weekends and weekdays. Any leak/burst events occurring on weekdays are therefore compared with previous weekdays and leak/burst events on weekends are only compared with previous weekends. This is because the night-time demand is often different on weekends due to the different consumption habits of customers during weekends, assuming that there are some residential customers as opposed to just a single large industrial metered consumer. To assemble the previous day data set, if the event occurs on a weekday, then the weekends are removed and more previous days are added to the data set so that the number of previous days is maintained. Determining the pressure residuals is then a matter of determining the leak/burst pressures and the normal pressures for each sensor location.

4.4.1.3. Normalising the Pressure Residuals

During the development of the sensor placement technique a procedure was developed for normalisation the changes in pressure generated by a hydraulic model which addressed several issues which are listed in bullet points below.

- The leak/burst localisation technique requires that the largest measured change in pressure should be in the vicinity of the leak/burst event.
- Sensor locations which are further from the inlet are often affected by leak/burst events which are not located in their vicinity. In many cases this means that events near to the inlet of the DMA will not successfully localised when there are sensors placed further downstream of the event.
- Even when the leak/burst events were successfully localised the search area will often be very large as it will include the sensor near to the inlet (and the leak/burst event) and the sensor(s) which are much further downstream.
- The typical response for a given sensor location to all leak/burst events is not accounted for.

The developed normalisation procedure overcomes these problems by comparing the measured changes in pressure, for each sensor, to the average change in pressure determined for many leak/burst events. To normalise the changes in pressure for each sensor location, normalisation factors derived from the hydraulic model data are used. The hydraulic model is used for two important reasons. Firstly, it allows the change in pressure for many leak/burst events to be determined without waiting for them to occur in the real DMA. Also, the changes in pressure in a real DMA are removed from consideration meaning that the change solely due to the leak/burst events is quantified. To do this, for each of the sensor locations, the average of the sensitivities for all of the leak/burst event sizes which were considered to determine the optimal sensor configuration are calculated. The changes in pressure for a new leak/burst event are then divided by the normalisation factors to give the normalised changes in pressure. The normalised changes in pressure are used to determine the approximate leak/burst location by the leak/burst localisation technique.

Further explanation of the normalisation process is given in section 4.4.2.3.3 which outlines how the normalisation is used by the sensor placement technique to analyse hydraulic model data rather than real leak/burst event data. The evidence which was gathered during development of the sensor placement is provided in that section.

4.4.1.4. Inverse-Distance Weighted Interpolation

Due to the limit on the number of sensors which can feasibly be deployed in a real WDS, which is driven by tightening regulatory targets and the budgetary constraints which are placed on water companies, techniques which can combine a small number of measurements to build up a wider picture of the behaviour in WDSs are of interest. Spatial analysis techniques have been previously proposed for the purposes of leak/burst localisation (Romano et al., 2013) and joint sensor placement and leak/burst localisation (Soldevila et al., 2019) in WDSs. The ability of these techniques to estimate the value of a variable at locations which are not measured means that they offer the potential to reduce the number of sensors required to estimate the behaviour of the entire DMA.

The heart of the leak/burst localisation technique is a novel version of IDW interpolation (Zimmerman et al., 1999). IDW interpolation is one of the most popular spatial techniques across a wide range of fields due to its simplicity, ease of computation and the need to only specify one parameter. As with many spatial analysis techniques, the estimated value at unmeasured locations is determined as a weighted sum of the measured values where the weight for each measured location is based upon its proximity to the location being estimated. The individual weight between one measured location, $\boldsymbol{i}$, and the unmeasured location, $\boldsymbol{j}$, is given by Equation 4-1.

$$\boldsymbol{W_{ij} = \frac{1}{d_{ij}^{n}}} \qquad \textbf{(\textit{Equation 4-1})}$$

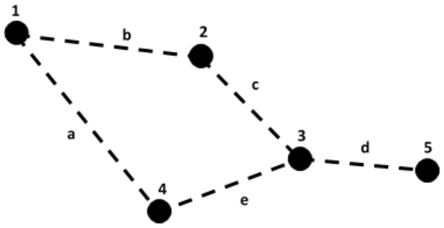
The weight between the two points is inversely proportional to the distance between the two points raised to the power of an exponent, $\boldsymbol{n}$**.** The distances between the measured locations and the unmeasured location are calculated using Euclidean distance, which is shortest straight line between the two points. The exponent is the single parameter which must be selected to perform the IDW interpolation technique. The value of the variable at the unmeasured location is calculated using Equation 4-2.

$$\boldsymbol{z_j = \frac{\sum W z_i}{\sum W}} \qquad \textbf{(}\textbf{\textit{Equation 4-2}}\textbf{)}$$

In Equation 4-2, the estimated value at the unmeasured location, $\boldsymbol{z_j}$, is the sum of the weights multiplied by the measured values which is normalised using the sum of the weights. This process of determining the distances, calculating the weights and then estimating the value at the unmeasured location is repeated for all unmeasured locations. Further details regarding selecting the best value for the interpolation exponent is given in section 4.4.2.4.6.

4.4.1.5. Spatially Constrained Distance Function

The novel step, which was introduced in Boatwright et al. (2016), was to modify the calculation of the distances. The reason behind this is that WDSs are constrained to the network of pipes and using the Euclidean distance in this case would not respect the distances that the water has to flow. The distances used by the novel, SC-IDW interpolation technique are determined as the shortest distance travelled between two points in the DMA being considered. The starting point to determine the distances is an undirected graph which is derived from a hydraulic model. Hydraulic models are often created from geographical information systems (GIS) data which ensures that the lengths of the pipes in the hydraulic model are representative of the pipes in the real WDS. There are several steps to processing a hydraulic model so that a graph can be constructed in order to perform the leak/burst localisation technique. The graph is constructed using a table of start and end nodes for each link and a corresponding weight for each link which is equal to its length. A simple example of a graph and the required list of start points, end points and link weights are shown in Figure 4-5. Each of the processing steps for converting the hydraulic model to a graph is shown in Figure 4-6.



Link	End Node 1	End Node 2	Weight
a	1	4	5
b	1	2	4
c	2	3	3
d	3	5	3
e	3	4	3.5

Figure 4-6 *Example undirected graph (L) with the link data table (R) used to create the graph*

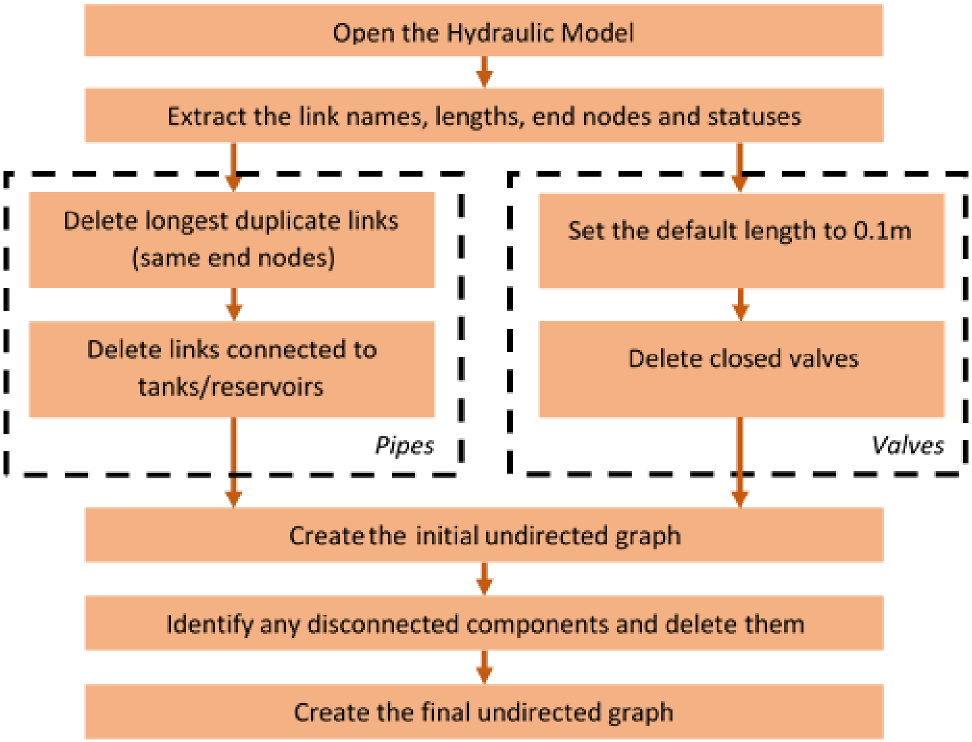


Figure 4-5 *The automatic process for creating a graph from a hydraulic model*

The entire process of creating the graph from the hydraulic model is automatic. The first processing step is to extract all of the data relating to the links in the DMA hydraulic model. The pipes and valves (and pumps if they are present) are considered separately. The longer of any set of links with the same end nodes is deleted as it is a condition of creating a graph in Matlab that duplicate links must be removed. To ensure that the shortest paths are always calculated the longest of the duplicate links is removed. Any links directly connected to either tanks or reservoirs are also removed as the interpolation does not need to estimate the pressures at the tanks or reservoirs.

For valves an assumed length of 0.1m has been used as a default value as this is not a property which is contained within the hydraulic model. Any valves which are closed in the hydraulic model are deleted as they cannot be traversed in the DMA and the graph should also reflect this. Once the valves and pipes have been processed an initial graph is created. At this stage there is the possibility that some links or nodes (which are used to connect the links) have become disconnected from the model. This usually occurs when a link is deleted. Any links or nodes which have been disconnected by previous processing steps are identified and removed so that only a single connected graph remains.

To perform the SC-IDW interpolation technique, using the distances travelled along the pipes rather than Euclidean distance, a matrix of the shortest distances between all pairs of nodes is generated. The matrix of nodal distances is a two-dimensional square matrix whose dimensions are both equal to the number of nodes in the graph. The matrix of nodal distances is symmetrical as the distance between two nodes is the same regardless of the direction of travel. A graph-based shortest path algorithm, called Dijkstra's algorithm (Dijkstra, 1959), is used to determine the shortest paths. The shortest paths between any two points are extracted from the nodal distance matrix and inserted into Equation 4-1 to determine the weights used by the SC-IDW interpolation.

4.4.1.6. Interpolation Exponent

One of the advantages of using IDW interpolation, as opposed to more complex spatial analysis techniques, is that only one parameter needs to be specified. This is the exponent used to calculate the weights between nodes, which depends upon the distance between them. The effect of the exponent is to control the amount of influence that near measured locations have when compared to distant measured locations on the value at the unmeasured point which is being estimated. The exponent influences the shape of the surface which is produced and using higher values of exponent tends to have a smoothing effect on the interpolation surface. Different values for the interpolation exponent have been used depending upon the application.

For example, de Mesnard (2011) used a value of 1 when considering the concentration of a pollutant in a vertically expanding cloud. However, when an explosion was introduced, which also increased the distance which needs to be considered an interpolation exponent of 3 was preferred. In both cases presented by de Mesnard (2011) the underlying physical processes and study space were used to derive the most suitable value for the interpolation exponent. Greenberg et al. (2011) estimated the temperature distribution throughout a river delta using a fixed value of 1 for the interpolation exponent. The best value for the exponent can also be selected using a sensitivity analysis in order to maximise some performance metric. One approach to this is to use leave one out cross validation (Arlot and Celisse, 2010) whereby a single measured location is removed from consideration and the agreement between that measured location and its value estimated using the other measured locations is used to measure the error. This process is repeated, excluding single measured locations in turn, for all measured locations, using various different values of exponent so that the value of exponent which minimises the total error can be determined. For the case where the instrument locations are already fixed then leave one out cross validation is a legitimate approach. Robinson and Metternicht (2006) concurred with this and obtained the best results when the interpolation exponent was determined using cross validation. The problem is that the best interpolation exponent will depend upon the sensor configuration which, in turn, depends on the interpolation exponent. A process for automatically determining the best value of the interpolation exponent was developed to overcome this which is described in section 4.4.2.4.6.

4.4.1.7. Determining the Search Area

Once the SC-IDW interpolation technique has been performed, the next step is to determine which parts of the DMA should be searched by field teams. The area with the largest estimated change in pressure is most likely to be in proximity to the leak/burst event so a method of dividing the surface of estimated values of the change in pressure is required. To achieve this the search area threshold is used. The search area threshold is selected or determined as a proportion of the range of values on the interpolation surface determined for a leak/burst event and can take any value between 0 and 1. The numerical value of the threshold, x_{sa}, is calculated using the maximum and minimum estimated values on the interpolation surface as shown in Equation 4-3.

$$x_{sa} = z_{j,min} + threshold_{sa}(z_{j,max} - z_{j,min}) \qquad \textbf{\textit{(Equation 4-3)}}$$

Any locations on the interpolation surface with an estimated value which is above the numerical threshold is designated as part of the search area. This information is then passed on to field teams to aid in the search for the leak/burst event to reduce the amount of time taken. The value of the search area threshold affects the proportion of the DMA which is included in the search area. The amount of the DMA which is in the search area is a measure of the localisation performance for a leak/burst event. Therefore, choosing the best value for the search area threshold is an important step in ensuring that good localisation performance for the leak/burst localisation technique is achieved. The sensor placement technique automatically determines the best value for the search area threshold whilst determining the optimal sensor configurations as described in section 4.4.2.4.6.

4.4.2. Sensor Placement Technique

4.4.2.1. Overview

The sensor placement technique makes use of the leak/burst localisation technique to determine the best locations of pressure sensors in a DMA so that the pressure sensors will be optimally located with respect to the leak/burst localisation technique as well as the DMA specific response to leak/burst events. Crucially, the localisation accuracy determined by the leak/burst localisation technique for many leak/burst events is minimised directly by the sensor placement technique. An automated procedure for determining the range of leak/burst event sizes has been developed in conjunction with an optimisation-based grouping procedure which significantly reduces the number of leak/burst event scenarios which are considered. The optimal sensor configurations are determined with respect to multiple sizes of leak/burst event because of this. The first step of the sensor placement technique is to model a range of leak/burst event sizes at various leak/burst event locations throughout the DMA to build a sensitivity matrix. The sensitivity matrix contains the changes in pressure for all potential sensor locations for all of the leak/burst event scenarios, where each leak/burst scenario is a leak/burst event size modelled at a single leak/burst event location. For each sensor configuration, the pressures corresponding to the selected sensor locations are extracted from the sensitivity matrix and analysed by the leak/burst localisation technique. For each leak/burst event the size of the search area is determined and the average size of the search area is calculated for the sensor configuration, considering all leak/burst events. For any event which is not correctly localised (i.e. the actual modelled leak/burst event location does not fall within the search area highlighted by the leak/burst event localisation technique) the whole size of the DMA is counted to assign a poor level of localisation for that leak/burst event. The average size of the search area for all leak/burst events (including the penalty for incorrectly localising leak/burst events) is used to measure the quality of each sensor configuration. The GALAXY multi-objective evolutionary algorithm (MOEA) (Wang et al., 2017) is used to search for the optimal sensor configurations in an efficient way due to the potential number of configurations which is possible for DMAs of normal size, mostly ranging between 500-5000 properties. The objective function used by the GALAXY MOEA (as part of the sensor placement technique) is formulated to minimise the average size of the search area and the number of sensors at the same time so that the optimal solutions with different numbers of sensors can be determined. The Pareto front of optimal sensor configurations can then be passed to water company decision makers to determine which optimal sensor configuration should be deployed for a given DMA.

4.4.2.2. Hydraulic Model Processing

4.4.2.2.1. Overview

The framework of methods requires a hydraulic model to determine the optimal sensor locations for a DMA. The process for converting and checking hydraulic model models is detailed in Figure 4-7. Hydraulic models are available for most DMAs in United Utilities WDS and this is likely to be the case for most water companies in developed countries. It is an assumption of the framework of methods which has been developed that a hydraulic model of sufficient quality (although this has not been defined) is available. Where this is not the case the sensor placement technique cannot be run.

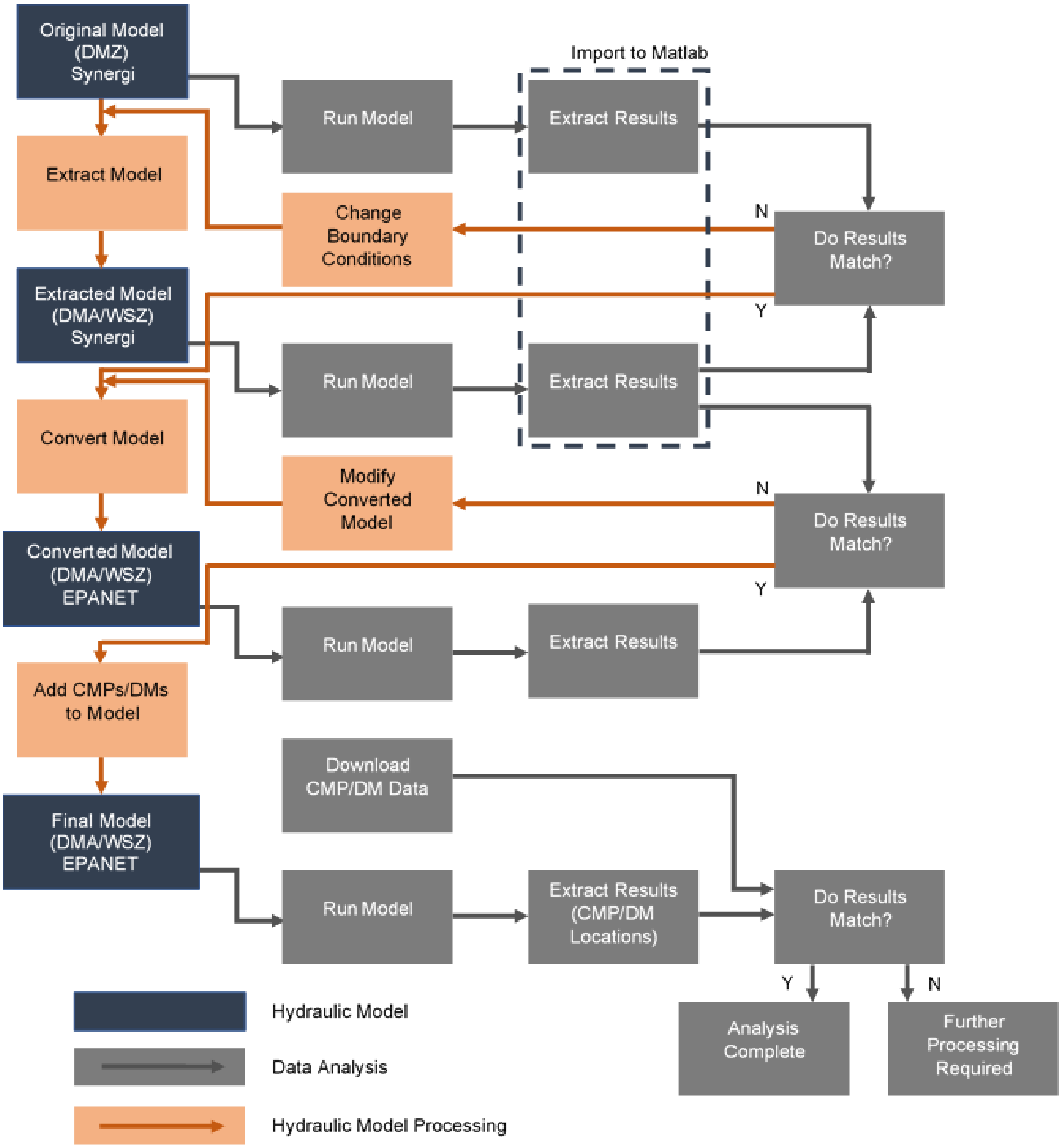


Figure 4-7 *Hydraulic model processing workflow for converting from Synergi to EPANET*

Hydraulic models at United Utilities are stored in the Synergi hydraulic modelling software package (DNV.GL, 2016). In order to be able to manipulate and analyse the models using the Matlab-EPANET wrapper (Uber and Hatchett, 2013) they need to be converted to a file type recognised by EPANET. Synergi has a feature which allows the models to be converted to a text file which can then be analysed using the Matlab-EPANET wrapper. There are several steps in extracting, converting and evaluating the models which are described in further detail in subsequent sections. This process is not fully automated due to the functionality for automatically exporting hydraulic models not existing in the Synergi hydraulic modelling software package. The process only needs to be completed once per DMA prior to running the sensor placement technique.

4.4.2.2.2. Hydraulic Model Extraction and Conversion

Hydraulic models at United Utilities are stored as an entire demand monitoring zone (DMZ) within the Synergi hydraulic modelling software package and typically contain between 10 and 50 DMAs. In addition to the DMAs, supplies such as those from service reservoirs are also contained within these models. This allows the connectivity of the system from the service reservoir to the customers to be viewed and the hydraulics analysed by hydraulic modelling staff at a water company, rather than viewing each DMA in isolation. The framework of methods developed in this project currently only needs to analyse each DMA individually and it is therefore of interest to be able to obtain a model of a DMA rather than having to analyse an entire DMZ model. There are several advantages to this which are listed below:

- Only the components contained within the DMA will be included in the hydraulic model. This reduces the chance of accidentally analysing the pressures from nodes outside of the DMA of interest and makes the job of generating the sensitivity matrix a far simpler task which can be checked more easily.

- The time to compute the results from the hydraulic model will be reduced. When only a few simulations are run this is not a problem but the sensor placement technique needs to consider many scenarios and therefore running the smallest possible model will be computationally efficient.

- The time to extract and convert the models is also reduced significantly, although given that this task only needs to be completed once for each DMA this most likely going to be small compared to the time savings which are made during the hydraulic model simulations.

The DMAs within the overall DMZ models are stored as subsystems in Synergi. Subsystems are entities within the overall hydraulic model for which a valid set of mass balance equations can be derived and solved. This means that each DMA can be extracted from the DMA model and still be solved in isolation. Synergi has a function which allows subsystems to be extracted into a standalone hydraulic model such that the boundary conditions are maintained. This means that the extracted model will, assuming that no errors have been made during the extraction process, run as though it were still contained within the overall hydraulic model. It is proposed to utilise this functionality to create a standalone DMA model for each DMA which is being analysed. A final step was added to check the standalone model against the original to ensure that the extraction process did not introduce any significant errors. Several numerical techniques are employed by Synergi, such as the Newton-Raphson method (Stroud and Booth, 2011) for solving the set of non-linear equations generated by applying Kirchoff's first law (Alexander and Sadiku, 2013) to the model in question. Further information about the computational methods which are used by Synergi can be found in the Synergi user guide (DNV.GL, 2016). When a model is extracted the set of equations changes and therefore the solutions may not be identical. However, the solutions should be very similar and the magnitude of the differences between the nodal pressures and the link flows for all time steps should be very small. The check is performed by extracting all the pressures and flows from Synergi along with the tags for all objects for the overall model and the extracted model.

Once the extracted model has been checked it can be converted so it can be recognised by EPANET, so that it can be analysed programmatically using the Matlab wrapper for EPANET (Uber and Hatchett, 2013). Synergi also supply a model conversion tool with their hydraulic modelling software package for this purpose (DNV.GL, 2016). During the conversion process each of the components stored in the Synergi model are converted to the equivalent type (or the nearest available) in

EPANET. There are some key differences between the components and their settings which are covered in the Synergi user guide (DNV.GL, 2016). These differences have the potential to cause different results in the two hydraulic modelling software packages which means that a secondary check is required to ensure that the model behaviour is maintained during the conversion process. The same procedure is repeated as for the hydraulic model extraction step except that the overall model data is replaced with the data collected from the EPANET model being run using Matlab. If there are any significant deviations between the two sets of model results then an investigation into the likely source is required and the model will need to be modified manually until it matches the extracted Synergi model outputs. These results are then compared in Matlab by calculating the differences across all time steps.

4.4.2.2.3. Evaluation of Hydraulic Model Quality

Once the model has been converted without any substantial errors then the last step is to evaluate the model against real data to determine how well calibrated it is. The hydraulic models used in this thesis were used at the same calibration level that is maintained by a water company. Further hydraulic model calibration fell outside of the scope of this project and there are many techniques for calibrating hydraulic models which could be used, if required, when the framework of methods is used operationally. The implications of the lack of hydraulic model calibration are explored further. Within the framework of methods, this evaluation step was being used to ensure the suitability of the DMAs which were used in the case studies which are presented later. In order to evaluate a model, data needs to be collected from the flow and pressure sensors which are deployed in the DMA being considered. At a minimum there will be flow data, collected every 15 minutes at all import and export points in the DMA, for all DMAs and one or two pressure readings collected from the critical monitoring point every 15 minutes. Using only one or two pressure measurements to evaluate the model accuracy is not ideal and it is desirable to have a larger number of pressure measurements to properly verify the model accuracy. Generally, the higher the number of pressure sensors available in a DMA the higher the confidence in the model's accuracy will be, assuming that all pressure measurements agree well with the results from the hydraulic model.

This means that the locations of the pressure and flow sensors in a DMA need to be verified. Any instruments which are not included in the hydraulic model will need to be added. For pressure sensors this can be achieved by adding a new node into the hydraulic model or using an existing node which is in the same location (or sufficiently close). An interesting question is how to set the correct elevation of the new/existing sensor nodes in the hydraulic model. Using the wrong elevation will mean that the pressures produced in the hydraulic model will be incorrect by an amount similar to the magnitude of the error in elevation. However, the approximated elevation of the installed sensor will have an associated error which means that the true value for both will not be known with certainty. Therefore, the proposed approach here is to use the elevation of the sensor node given in United Utilities' GIS database initially and then to offset this to improve the agreement if the pressures vary by an approximately constant amount for all time steps. If the differences between the pressures from the hydraulic model and the sensor data vary significantly (by more than the accuracy of the instruments being used) over the considered time steps then it is clear that something other than incorrect elevation at the pressure sensor locations is responsible for the error. As the analysis window has been set to consider only the night then the amount of variation in the pressure (and the differences between the model and the sensor data) is typically greatly reduced because the hydraulic model parameters, such as the demand profiles, will vary less.

In the case that for all pressure sensors the modelled pressures for all time steps are similar to the measured pressure data, then the model is considered accurate enough to be used for sensor placement. A qualitative assessment is performed so that the average difference between the model and the sensor data is less than 1m, determined during the analysis window. Determining how much difference between the modelled results and measured data is acceptable for good results to be achieved by the sensor placement technique is an important step but it is outside of the scope of the case studies provided here.

4.4.2.2.4. Pressure-Driven Demand

Pressure-driven demand refers to a group of techniques for modelling WDSs under conditions where there is insufficient pressure available to supply the demands. Within EPANET, when demands are modelled, it is assumed that they are fully satisfied regardless of how much pressure there is at the point at which the demand is modelled. This is not realistic and does not match the behaviour of demands in real WDSs. In developed countries there are normally regulations related to the minimum amount of pressure that must be provided to customers to ensure that they do not experience problems with their water supply. In the UK, this is covered by the guaranteed standards scheme (Ofwat, 2017) which dictates that the minimum pressure to be maintained at the customer stop tap is 0.7 bar. In practise, this is not measured directly but the common approach uses by water companies is to collect a small number of pressure measurements in a DMA at the point which is most sensitive to low pressures. Typically, this is either at the point of highest elevation or the point furthest from the inlet of the DMA. In some cases, multiple pressure instruments are used depending upon the size and complexity of the DMA. When an abnormal event occurs, for example a leak/burst event there can be widespread reductions in pressure which affect customer supplies. To model these situations accurately it is important to account for the pressure-driven behaviour of the demands.

The technique from the aforementioned proposals which has been incorporated into the framework of methods is the approach proposed by Paez et al. (2018). A brief summary of the implementation of this technique is now given. Firstly, to each demand node a combination of a flow control valve (FCV), throttle control valves (TCV), check valve (CV) and a reservoir are added to enforce Wagner's relationship (Wagner et al., 1988) between pressure and flow. The FCV is used to limit the demand so that its maximum value is the same as the original demand. The throttle control valve is used to control the pressure flow relationship by altering the amount of pressure which is lost for a given flow rate. The reservoir is added to maintain the static head which must be overcome for flow to occur and the check valve prevents the flow from reversing from the reservoir back into the DMA. The main benefit of using the approach by Paez et al. (2018) is that by using the TCV to enforce the pressure-flow relationship simple controls within the EPANET software package can be added to adjust the TCV for an extended period simulation rather than using the programmer's toolkit for the same purpose.

Two parameters need to be specified prior to implementing the pressure-driven demand approach by Paez et al. (2018). The parameters are the minimum pressure and required pressure. The minimum pressure is the pressure at which flow starts to occur and the required pressure is the pressure required to achieve the full demand. For the purposes of the framework of methods the minimum pressure has been set to 0m and the required pressure has been set to 7m to match the pressure required under the guaranteed standards scheme.

4.4.2.3. Leak/Burst Modelling Technique

4.4.2.3.1. Overview

The basis of many sensor placement techniques for leak/burst detection and localisation is modelling a range of leak/burst event scenarios so that the response of a DMA to these events can be used to determine the optimal set of sensor locations.

4.4.2.3.2. Sensitivity Analysis

Sensitivity analysis is well established in the sensor placement techniques which have been developed for placing flow and pressure sensors in WDSs. A sensitivity analysis determines, for all possible combinations of leak/burst locations and leak/burst event sizes, the changes to both the pressure and, in some cases flow, which will occur. The purpose of this is to capture the changes in pressure and flow at the potential sensor locations so that this data can be searched to identify the best set of sensor locations. The procedure for performing a sensitivity analysis is outlined in Figure 4-8.

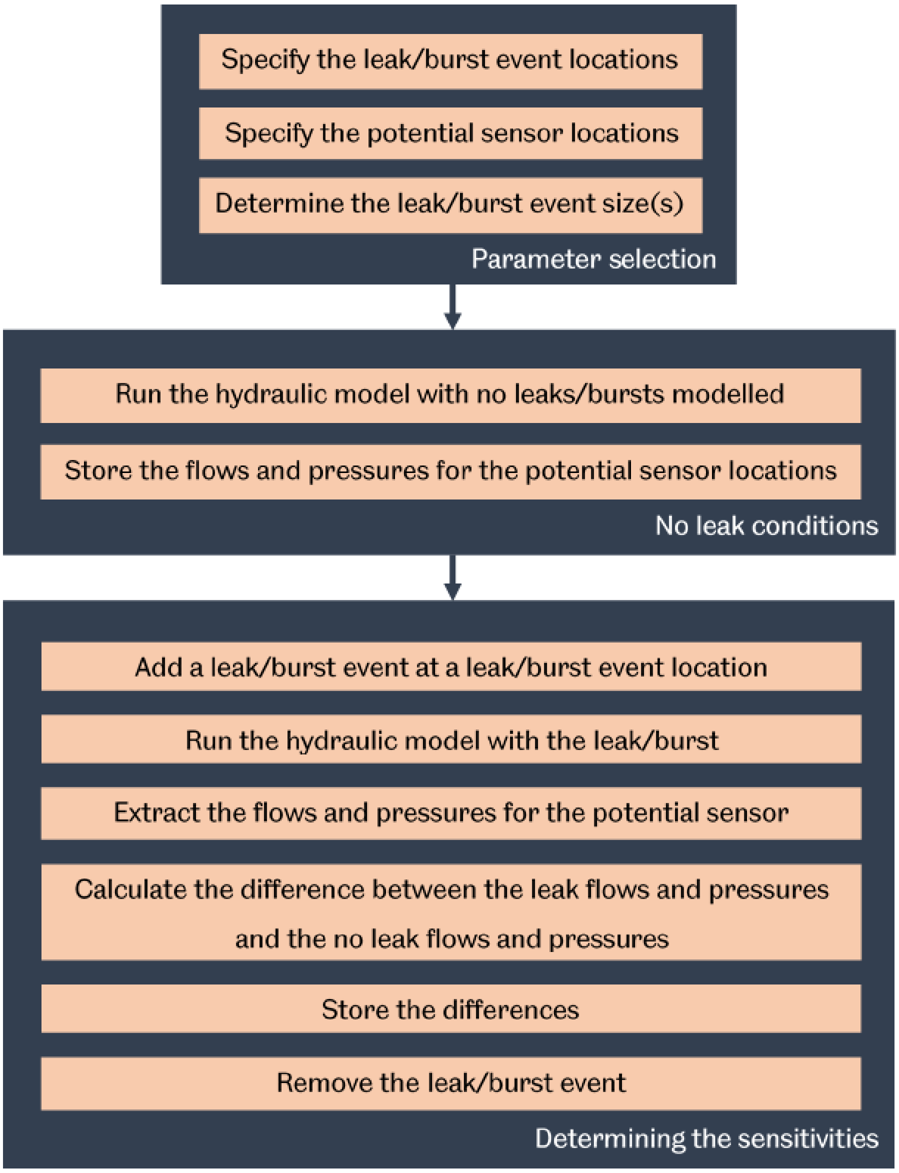
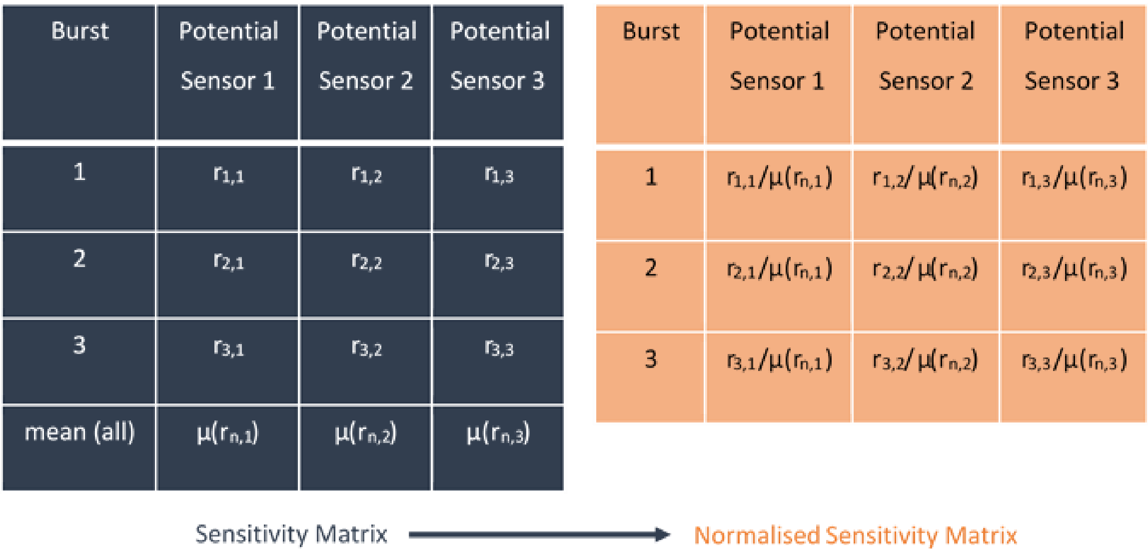


Figure 4-8 *Flowchart of the sensitivity analysis*

Several parameters must be specified or determined prior to performing the sensitivity analysis. The first two, which are related to the leak/burst events are the leak/burst event sizes and the leak/burst event locations. Additionally, one parameter related to the sensor locations must also be specified. This is the set of potential sensor locations. Only pressure data from these points needs to be considered and data from any other point in the hydraulic model are ignored during the sensitivity analysis. Next, a single run of the hydraulic model is performed so that the normal pressures and flows are captured. These will be used to determine how much the pressure and flow has changed for each leak/burst scenario. The final step is to iteratively add leak/burst events to the hydraulic model, one at a time, and storing the changes in pressure which occur as a result. The sensitivity matrix is the method for storing the changes in pressure and flow which arise as a result of leak/burst events. Typically, the changes in pressure and flow, for the potential sensor locations, for a single leak/burst are stored as a single row (or column) in a matrix. Subsequent rows are added for each new leak/burst event when multiple leak/burst event sizes are used a new sensitivity matrix is typically constructed for each leak/burst event size (or for each combination of different parameters which are being considered).

4.4.2.3.3. Normalising the sensitivity matrix and derivation of normalisation factors for multiple sensors

To prevent the information loss which is associated with binarizing the sensitivity matrix (Perez et al., 2009), a non-binarised sensitivity matrix is used as the input to the leak/burst localisation technique used by the sensor placement technique. This allows the magnitude of the change in pressure to be determined rather than using a binary value which is not compatible with the leak/burst localisation technique. The process for normalising one sensitivity matrix is shown in Figure 4-9.



Burst	Potential Sensor 1	Potential Sensor 2	Potential Sensor 3
1	$r_{1,1}$	$r_{1,2}$	$r_{1,3}$
2	$r_{2,1}$	$r_{2,2}$	$r_{2,3}$
3	$r_{3,1}$	$r_{3,2}$	$r_{3,3}$
mean (all)	$\mu(r_{n,1})$	$\mu(r_{n,2})$	$\mu(r_{n,3})$

Burst	Potential Sensor 1	Potential Sensor 2	Potential Sensor 3
1	$r_{1,1}/\mu(r_{n,1})$	$r_{1,2}/\mu(r_{n,2})$	$r_{1,3}/\mu(r_{n,3})$
2	$r_{2,1}/\mu(r_{n,1})$	$r_{2,2}/\mu(r_{n,2})$	$r_{2,3}/\mu(r_{n,3})$
3	$r_{3,1}/\mu(r_{n,1})$	$r_{3,2}/\mu(r_{n,2})$	$r_{3,3}/\mu(r_{n,3})$

Figure 4-9 *Normalisation of a sensitivity matrix*

The changes in pressure for each sensor location are divided by the normalisation factor, determined using a hydraulic model specifically for each potential sensor location. The normalisation has the effect of scaling the sensitivities by an amount which is proportional to the average sensitivity for a given potential sensor location to overcome the issues detailed in section 4.4.1.3. A simple hydraulic model with 3 sensors and 3 leak/burst event locations is shown in Figure 4-10. In Table 4-1 the sensitivities were calculated for each of the leak/burst event locations and measured at the three sensor locations. For leak/burst event 1, which is located close to the inlet, the two sensors located downstream at points 1 and 2 measured a slightly higher reduction in pressure than the sensor located close to the leak/burst event location. For leak/bursts events 2 and 3 the sensor nearest to each event also measured the largest change in pressure, which was expected. Considering this behaviour, it is clear that localising the leak/burst event at location 1 would be difficult using the leak/burst localisation technique. The normalisation step allows this problem to be resolved by accounting for the higher average sensitivity of the downstream sensor locations. There are several reasons why the sensitivities are different for different potential sensor locations. Firstly, potential sensor locations which are further from the inlet of the DMA will be affected by a greater number of leak/burst events. This is because the largest changes in pressure are normally seen downstream of the location of the leak/burst event. Secondly, because all of the junctions in the DMA are considered as valid leak/burst event locations, the distribution of the junctions in the DMA hydraulic model will also affect the apparent sensitivity of some potential sensor locations. Potential sensor locations which are close to many junctions will have a higher average change in pressure. However, it is undesirable to allow both of these factors to dictate which potential sensor locations are most sensitive and therefore, they both must be accounted for. To achieve this a normalisation step is performed on the sensitivity matrices prior to determining the optimal sensor configurations. The normalisation step modifies the values of the sensitivities which were determined by modelling many different leak/burst scenarios to account for both of the factors which can increase the apparent sensitivity of some potential sensor locations.

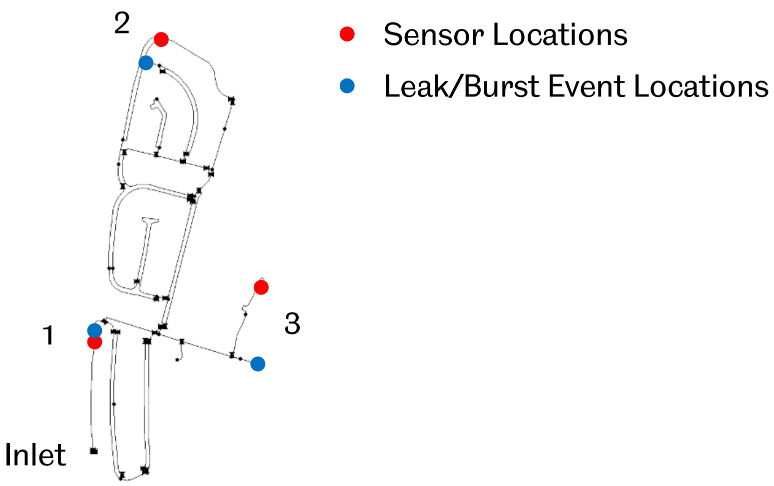


Figure 4-10 *Schematic of the simple hydraulic model with leak/burst event locations (blue) and sensor locations (red)*

Comparing the sensitivities in Table 4-1 with the normalised sensitivities in Table 4-2 shows how the normalisation of the sensitivities using the normalisation process shown in Figure 4-9 affects the sensitivities for the three leak/burst events in the simple hydraulic model. The key point is that the absolute magnitude of the changes in pressure are not the most important factor but the relative magnitudes when compared to other sensor locations. After the normalisation step was used the sensors in proximity to each of the leak/burst events had the highest values for the normalised sensitivities. This normalisation step and the method for deriving the normalisation factors is used by both the sensor placement technique and the leak/burst localisation technique to maintain consistency between them.

Table 4-1 *Sensitivities calculated for three leak/burst events at three sensor location using a simple hydraulic model*

Burst Location	**Pressure Sensor 1**	**Pressure Sensor 2**	**Pressure Sensor 3**
1	0.3298	0.3299	0.3299
2	0.3375	0.3432	0.3415
3	0.3883	0.3929	0.4112

Table 4-2 *Normalised sensitivities calculated for three leak/burst events at three sensor locations using a simple hydraulic model*

Burst Location	**Pressure Sensor 1**	**Pressure Sensor 2**	**Pressure Sensor 3**
1	0.937	0.929	0.914
2	0.959	0.966	0.946
3	1.104	1.106	1.139

4.4.2.3.4. Selection of Potential Sensor Locations

In order to determine the sensor configurations a set of potential sensor locations needs to be defined. There are several considerations which affect which locations in a DMA can be considered suitable for installing sensors in a DMA.

- **Cost of installation:** The cost of installing a single sensor is a function of its location. Installing a new sensor on a buried pipe will require the pipe to be exposed and the complexity of this task is related to the material in which the pipe is buried. For example, a pipe buried in a grass verge will be easier and quicker to expose than one buried in a road.
- **Ease of access for installation/maintenance/repair/replacement:** A primary concern for any asset owned by a water company is that it can be easily accessed by operations staff so that, in the event of a fault, equipment failure or routine maintenance such as a battery replacement it can be accessed. Access to some locations may be restricted if, for example, the pipe is located on private land.
- **Availability of power:** Many sensors on the market come supplied with batteries which last for several years meaning there is no need for a power supply to be installed when the sensor is installed. However, for some sensors, such as high frequency loggers, the battery life may be limited and a power supply may be required.
- **Water company engineering standards and/or preferences:** Large water companies in the UK have design and construction standards which must be

adhered to. For example, some companies may specify that new underground equipment must be installed in a chamber which has a significant impact upon where it is practical to install equipment.

- **National/International Standards and Legislation:** Additionally, in developed countries there are stringent laws governing such issues as healthy and safety, confined spaces, explosive atmospheres and many others. Each of these could be an issue which makes certain locations unsuitable for the installation of sensors in WDSs.
- **The types of sensors which are being used and how they can be installed:** The characteristics of the sensors being used will dictate where they can be installed. The physical characteristics such as the size of the sensor could rule out some potential locations.

Each of these factors must be considered to determine which locations can be considered as feasible or practical for installing sensors. Generally, many of these considerations have been largely ignored by previous sensor placement techniques. This is because much of the research has been targeted towards detection and/or localisation performance using a hydraulic model and, as such, most of these considerations are not relevant. However, when the real-world scenario is considered and sensors are deployed in a WDS they are very important. Many sensor placement techniques consider all junctions in the hydraulic model as potential sensor locations which can lead to difficulties when trying to deploy the optimal sensor configurations, determined using a hydraulic model, in real WDSs.

With all of these considerations in mind the optimal sensor placement technique considers fire hydrants as the set of potential sensor locations. This is to overcome many of the difficulties which have been highlighted. Fire hydrants are located in underground chambers which ensures that the sensors will be safe from vandalism or theft. The chambers can be accessed by water company operations staff so that they can be installed easily and any maintenance performed, if needed. There are several potential benefits and drawbacks of this approach, however. Firstly, a benefit is that using a smaller number of potential sensor locations will significantly reduce the size of the search space. This was noted by Farley et al., (2008), as they limited the number of pressure sensors to 2 which, in turn, meant that all possible sensor configurations could be searched using full enumeration. Another benefit of using fire hydrants is that their locations are already included in the hydraulic models available at United Utilities. This means that no further work is required to add the locations as junctions within the hydraulic model. Where this is not the case it is still a relatively simple, albeit time consuming, exercise to add the hydrants into the hydraulic model. A drawback is that by limiting the number of potential sensor locations the achievable performance of the sensor placement will be reduced. However, given that using other sensor locations would increase the cost of installing the sensors (due to the reasons already stated) and raise other issues, the potential loss of performance is considered acceptable.

4.4.2.3.5. Selection of the Analysis Window

In order to determine the changes in pressure an analysis window must be selected. The standard approach to hydraulic modelling is to use extended period simulation which considers the pressures and flows varying over multiple time steps. A typical hydraulic model used by water companies uses a 24-hour duration and a 15-minute time step where the time step denotes the amount of time which has lapsed in the hydraulic model between different calculation steps. The behaviour of a WDS is driven by the water consumption habits of customers and therefore the pressure and flows follow a diurnal pattern. The greatest difference between the normal pressure in a WDS, without the leak/burst present, and the pressure without the leak/burst present will occur during the night because the consumption by customers is lowest. This increases the pressure which, in turn, causes the leak/burst flow to be at its maximum. This fact has been exploited by water companies when they use MNF analysis to detect leak/burst events. In relation to leak/burst event detection and localisation analysing the night-time pressure offers the most promising window for capturing the changes in pressure. This is because, for a fixed minimum measurable change in pressure smaller leak/burst events will produce changes in pressure which can be measured. This means that smaller leak/burst events can be localised using additional pressure instrumentation. The analysis window which will be used to localise leak/burst events was set to consider the pressures between 03:00 and 04:00 which coincides with the period of minimum flow in most DMAs. It must be noted that since the analysis window is being considered over multiple successive days that the earliest that a leak/burst event can be localised is at 04:00 after the leak/burst event has occurred. This means that a maximum period of 24 hours can elapse between the leak/burst occurring and it being localised. However, given that the types of leak/burst events being targeted by the framework of methods are unreported, and therefore not likely to impact upon customer supplies, this is not seen as a serious limitation. In addition to this, for leak/burst events which are not impacting customer supplies some time may be taken for a water company to decide upon the best strategy for repairing a leak/burst event. Often this time can be longer than 24 hours which means that no extra time is incurred in the leak/burst repair time.

4.4.2.3.6. Automatic Determination of Leak/Burst Event Sizes

A variety of approaches have been used to select the event size or range of event sizes which were considered by previous sensor placement techniques. These range from considering a single fixed size of base demand at all leak/burst event locations (Perez et al., 2009), considering a single emitter coefficient at all leak/burst locations (Farley et al., 2008). Some sensor placement techniques (Casillas et al., 2013, 2015) even consider multiple emitter coefficients at each leak/burst event location. The approach used by Casillas et al. (2013, 2015) is better because in real WDSs leak/burst events of different sizes can occur and any deployed sensor configuration should be able to localise leak/burst events of multiple sizes. However, determining the correct size or sizes of leak/burst events to model represents a significant task because there are many important considerations.

The first consideration is that water companies only have a limited amount of resources with which to perform leak/burst localisation. Different sizes of leaks/burst can be localised using different tools and techniques and the optimal strategy should aim to minimise the total amount of resources which are used to obtain adequate leak/burst localisation performance. Larger events will typically have a larger effect on the hydraulics of a WDS and will also affect the supplies of customers. Therefore, it is highly likely that the presence and location of these leak/burst events will be readily apparent to a water company using current techniques or strategies. On the other end of the scale of leak/burst event size some events are so small that no strategy currently employed by water companies can be used to detect or localise them. This is primarily because the effects that they cause, such as increased flow, reduced pressure, or noise/vibration are imperceptible from the normal variations which occur in WDSs. Another contributing factor is that the cost associated with employing a strategy which could detect or localise them would actually be detrimental to the performance of a water company. The size of a leak/burst event also dictates how detrimental it is to the performance of a water company. Large events typically impact the supplies of customers and must be dealt with swiftly in order to reduce the impact on targets related to customer supplies. Smaller events, by virtue of the difficulty in detecting and localising them, will lead to large volumes of water loss which impacts upon regulatory targets and costs a water company as they must increase their production of treated water to compensate for the volume lost.

The key point is that there are a range of intermediate event sizes which are both too small to be readily apparent but also large enough to be localised using additional instruments placed throughout a DMA. It is possible to use a hydraulic model to determine the minimum size of leak/burst event which can be detected and localised for each leak/burst location. Assuming a complete sensor configuration, where all potential sensor locations have a sensor installed, then if any of the sensors register a change in pressure which is higher than the minimum amount of change which can be captured by that instrument then the event can be detected. In this case, the differential responses of all sensors can be used to infer the location of the leak/burst event. The minimum amount of change that can be captured by a sensor, which is also called the sensors accuracy, is a property of the instrument which needs to be accounted for when the leak/burst event sizes are determined. This is overlooked by many sensor placement techniques. The novel procedure for automatically determining the range of leak/burst events is shown in Figure 4-11.

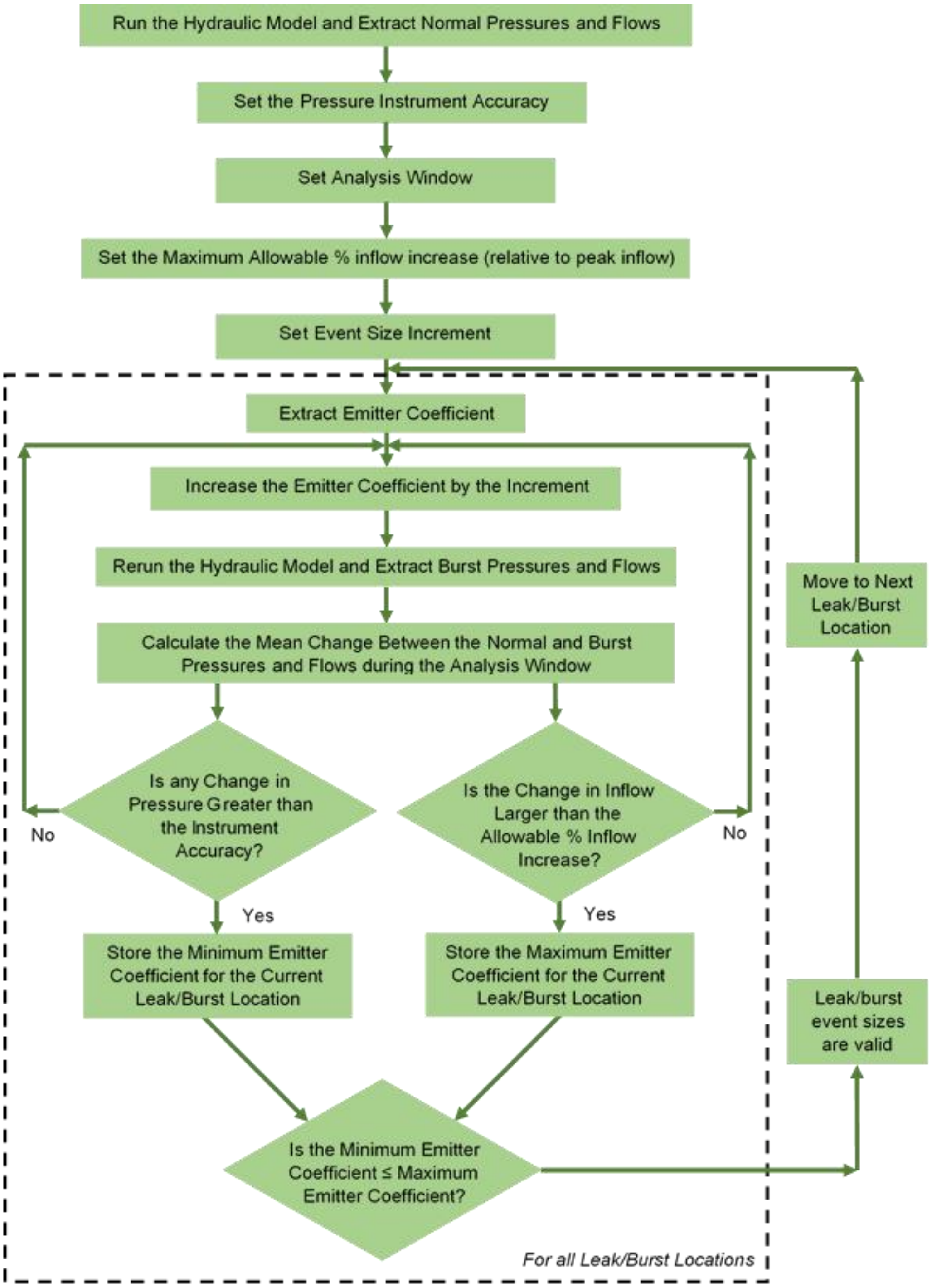


Figure 4-11 *Determining the minimum and maximum leak/burst event sizes*

The basic approach for determining the minimum leak/burst event sizes is to, for each leak/burst event location, incrementally increase the leak/burst event size until one potential sensor location registers a change in pressure which is greater than or equal to the instrument accuracy. This ensures that only leak/event sizes which can be measured by at least one sensor are considered for each leak/burst event size. It also allows different leak/burst sizes or range of leak/burst event sizes to be determined for each leak/burst event location. This is fundamental, as the basis for sensitivity analysis is that there exists a difference between the sensitivity of some leak/burst event locations. However, not all sensor placement techniques use different minimum leak/burst event sizes for each leak/burst location. This means that some leak/burst event scenarios are considered which do not produce large enough changes in pressure to be measured by pressure instruments.

To determine the minimum leak/burst event size for a single leak/burst event location a leak/burst event size (or emitter coefficient) of zero is set as the initial value. The hydraulic model is run and the pressures are extracted for all potential sensor locations for time steps coinciding with the chosen analysis window. These pressures, which relate to the normal behaviour of the model, are stored. Then the event size is systematically increased by a fixed amount and the pressures for all potential sensor locations are extracted in the same manner as the normal pressures. For each leak/burst event size and for all potential sensor locations the changes between the normal pressures and the leak/burst pressures are calculated. If any of the changes are greater than the instrument accuracy then the leak/burst event size is stored and the next leak/burst event location is considered. This process is repeated for all leak/burst locations.

A similar process is used to determine the maximum leak/burst event sizes. However, the leak/burst event sizes are not calculated with respect to the change in pressure but the increase in flow which occurs as a result of each leak/burst event. To achieve this an additional parameter is used which is called the maximum allowable flow increase. This parameter is a percentage which relates the average daily flow to the additional flow which has occurred as a result of a modelled leak/burst event considered by the sensor placement technique, calculated during the analysis window. The value of this parameter must be set prior to determining the maximum leak/burst event sizes and can be set using knowledge of the individual DMA in question or a fixed value can be used across an entire WDS to align with the size of leak/burst events which are to be targeted. Once the maximum allowable flow increase has been chosen/determined for a DMA the same iterative process of evaluating each leak/burst location is used as for determining the minimum leak/burst event sizes. Further details of how to determine the value of the maximum allowable flow increase are given in Case Study 5.8 in section 5.4.5.

Once the maximum and minimum leak/burst event sizes have been determined for all leak/burst locations, and the validity of the leak/burst event sizes are checked, the maximum and minimum leak/burst event size for the entire DMA can be determined. The maximum size for the DMA is the greatest of the individual maximum leak/burst event sizes and the minimum leak/burst event size is the smallest of the individual minimum leak/burst event sizes. The final leak/burst event sizes which are considered from this point forward are between the maximum and minimum leak/burst event sizes for the DMA. The sizes are determined by starting at the minimum leak/burst event size for the DMA and increasing the event size by the same increment that was used to determine the individual maximum and minimum leak/burst event sizes until the maximum leak/burst event size for the DMA is reached. For each of the final leak/burst event sizes a sensitivity matrix is generated in the same way as was outlined previously.

For the purposes of the case studies provided in this thesis a default value of 0.1 for the maximum allowable flow increase, equivalent to 10% of the daily peak flow, has been used. The event size increment must then be selected to ensure that the number of leak/burst event considered by the leak/burst event location grouping technique is reasonable. For some DMAs, particularly those that are not sensitive to leak/burst events, only large values for the maximum allowable increase in flow will produce valid results. This provides an indication that the DMA in question is not a good candidate for deploying pressure sensors because only large leak/burst event sizes cause sufficient change in pressure to be measured by any sensor configuration.

4.4.2.3.7. Leak/Burst Event Location Grouping Technique

Once the leak/burst event sizes have been determined another novel step in the framework of methods is to group together the leak/burst event scenarios according to the sensitivities. This is an important step because if multiple leak/burst event sizes are considered then the number of leak/burst event scenarios which need to be considered by the sensor placement technique increases proportionally to the number of leak/burst event sizes. The purpose of the leak/burst event grouping technique is to reduce the number of leak/burst events which are considered by the sensor placement technique by grouping together those leak/burst event scenarios which are similar in terms of the sensitivities at all possible sensor locations. The overall process of determining the grouped sensitivity matrix is shown in Figure 4-12.

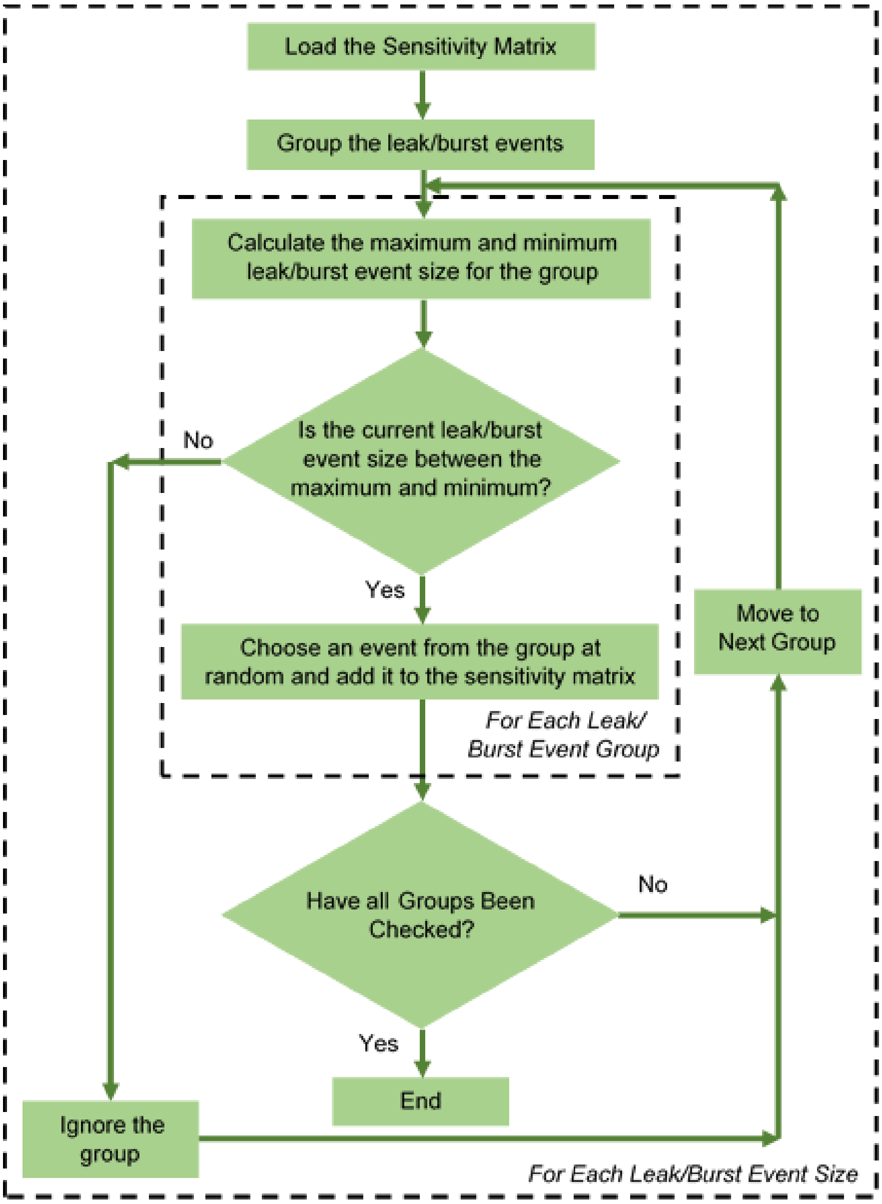


Figure 4-12 *Assembling the grouped sensitivity matrix for multiple leak/burst event sizes*

Previously, Sophocleous et al., (2019a) developed a model-based leak/burst localisation technique which grouped together the potential leak/burst event locations spatially by using a buffer zone around each leak/burst event location. In effect, any leak/burst events which were sufficiently close to each other were treated as the same leak/burst event. This is premised upon the notion that two leak/burst events separated by a small distance will behave in a similar way. The aim of this was to reduce the search space for a GA using the leak/burst locations as the decision variables. In relation to sensor placement the same logic can be applied for the purpose of reducing the number of events which must be considered for sensor placement.

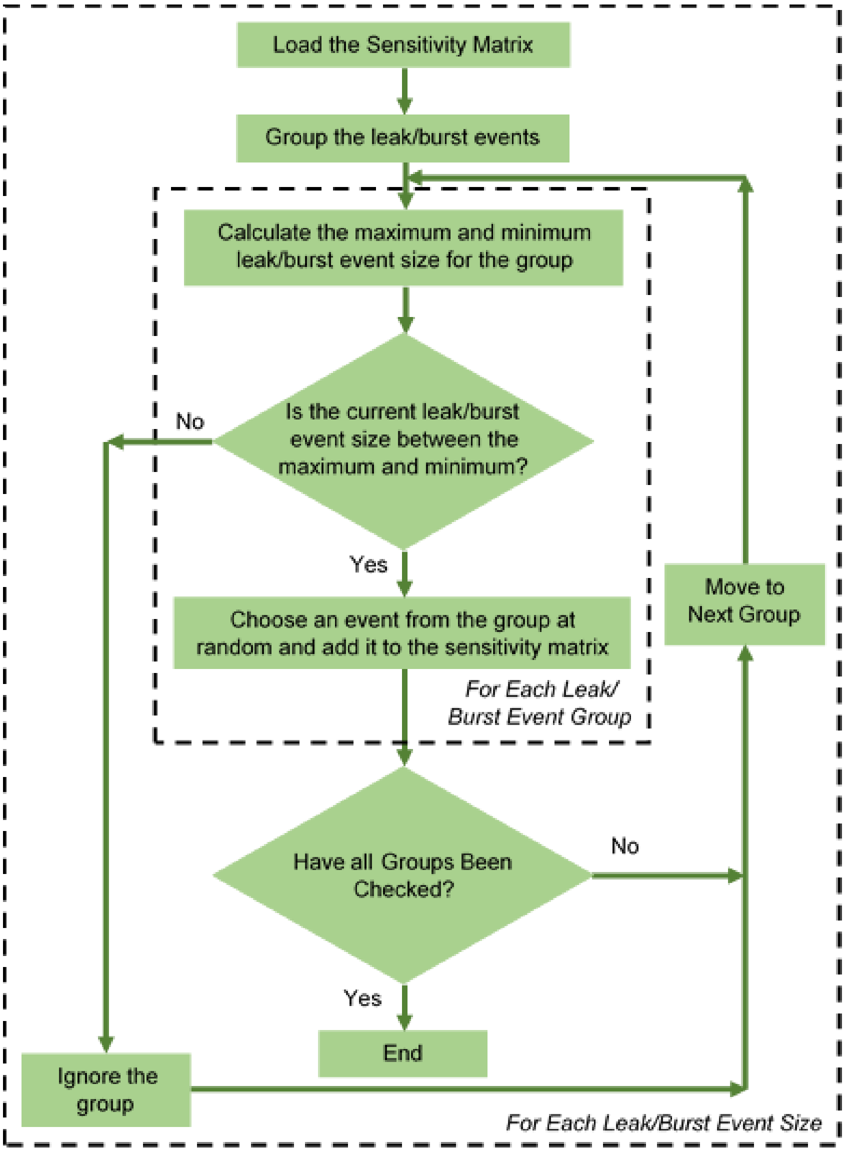


Figure 4-13 *Assembling the grouped sensitivity matrix for multiple leak/burst event sizes*

Considering that a typical DMA will contain several hundred leak/burst event locations, the number of leak/burst event scenarios when multiple leak/burst event sizes are used can be of the order 10^3-10^4. For the framework of methods presented here that means that the leak/burst localisation technique must be run 10^3-10^4 times for each considered sensor configuration. Therefore, any reduction in the number of leak/burst event scenarios which are considered is beneficial to the total computational effort (and time taken) to determine the optimal sensor configurations.

If each member of the set (in this case all leak/burst event locations) can only belong to a single subset then the number of possible groupings is equal to the Bell number (Bell, 1934), which is calculated using Equation 4-4. The Bell number describes the summation of the Stirling numbers of the second kind (Fort, 1948) over the entire range of possible numbers of groupings, k, which is between 1, where all members of the set are in the same subset, and n_{lb}, where the set is grouped into n_{lb} subsets each with a size of 1.

$$\boldsymbol{B_n} = \sum_{\boldsymbol{k}=1}^{\boldsymbol{n_{lb}}} \left\{ \begin{matrix} \boldsymbol{n_{lb}} \\ \boldsymbol{k} \end{matrix} \right\} \qquad \textbf{(\textit{Equation 4-4})}$$

By considering the problem of grouping the leak/burst event locations in an example DMA (DMA 123-09) the number of possible groups of the 425 leak/burst event locations is 9.67×10^{127}. This number only applies for unconstrained grouping but given that for grouping the leak/burst event locations the available hydraulic data will be used to determine which groups are valid. However, it is not known in advance which potential groups are valid and therefore the number of valid groups cannot be ascertained *a priori*.

Due to the complexity of solving this problem an optimisation-based approach to grouping the leak/burst event data has been proposed. The aim of grouping the leak/burst event locations is to use the minimum number of groups which satisfies the constraint related to the similarity of the members of each group. The minimum number of subsets is desirable as that will remove the largest number of leak/burst event scenarios. To achieve this GALAXY MOEA, which is also used to determine the optimal sensor configuration(s), has been utilised. The GALAXY MOEA requires integer decision variables so the problem will be formulated with the group indices for each leak/burst event location as the decision variables. Once the groups of leak/burst events have been generated each group must be checked to ensure it is valid. The process of checking the validity of a single group of leak/burst events is shown in Figure 4-13.

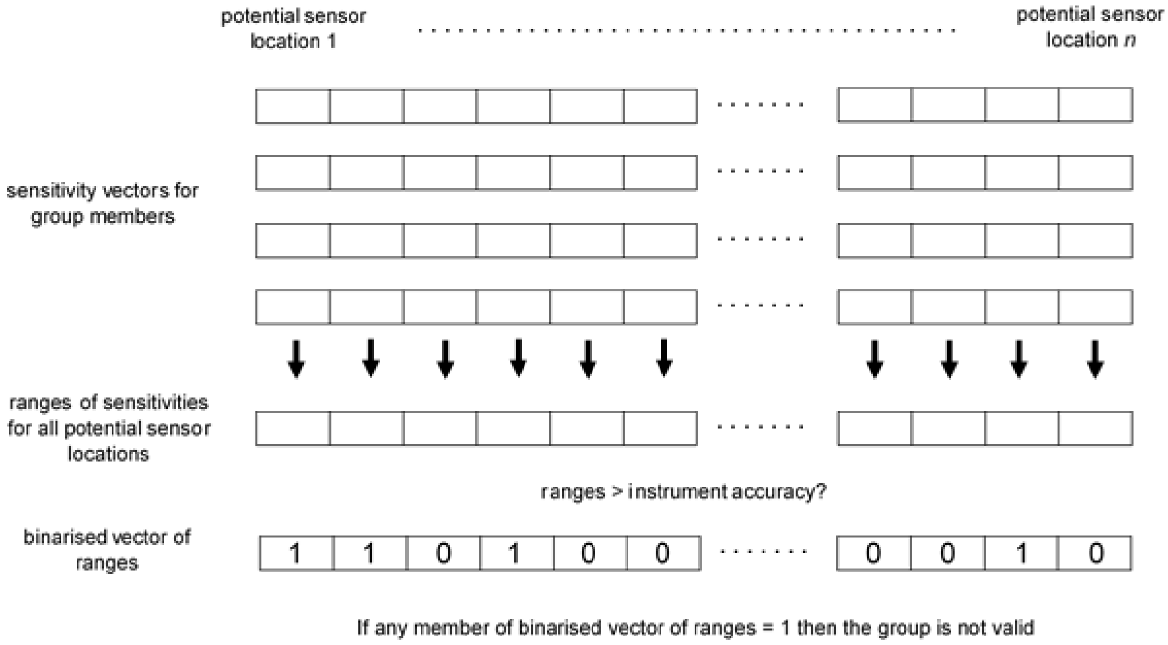


Figure 4-14 *Process of checking the validity of a single group of leak/burst events*

For each group of leak/burst events generated by the GALAXY MOEA, the vector of sensitivities corresponding to the group members is extracted. For each potential sensor location, the range of the sensitivities is checked to ensure it is below the instrument accuracy. If this holds then the group is valid. The procedure is repeated for all groups and if all groups are valid then the proposed solution is considered valid. This is demonstrated for a simple case with five leak/bust event locations formed into two groups in Figure 4-14.

The fitness of a set of groups, where all groups are valid, is equal to the number of groups. If any of the groups is not valid in a set of groups then the fitness is penalised to take the maximum possible number of groups, which is also equal to the number of leak/burst events. Using this process for checking the validity of the groups of leak/burst events the GALAXY MOEA can determine the minimum number of possible groupings which can be used for each leak/burst event size. The grouping must be performed for each leak/burst event size because the sensitivities vary as the leak/burst event size does.

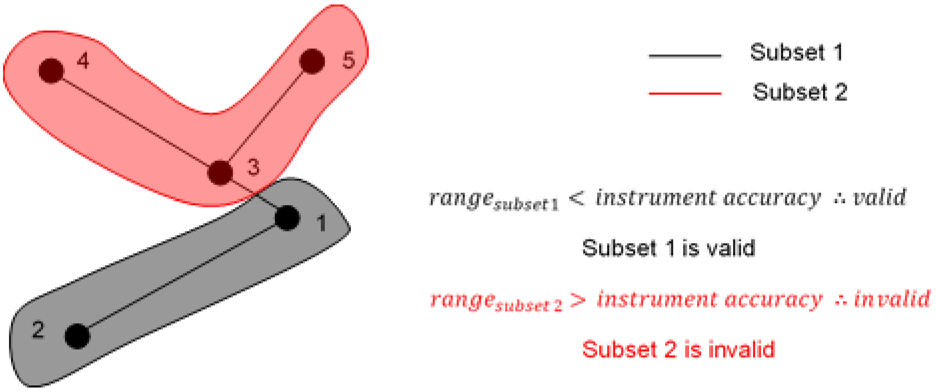
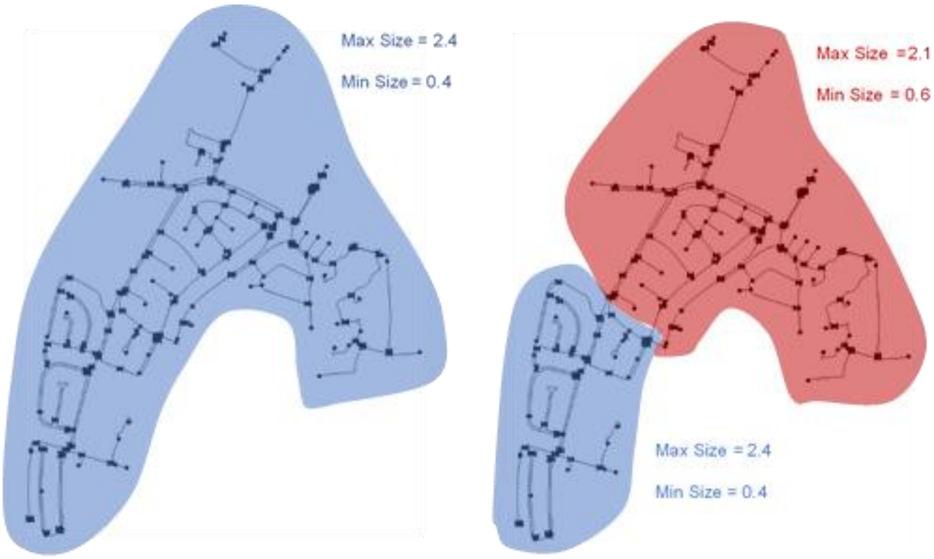


Figure 4-15 *Five leak/burst events formed into two subsets. Subset 1 is valid and Subset 2 is invalid. As one of the subsets is invalid the leak/burst events cannot be divided in this way*

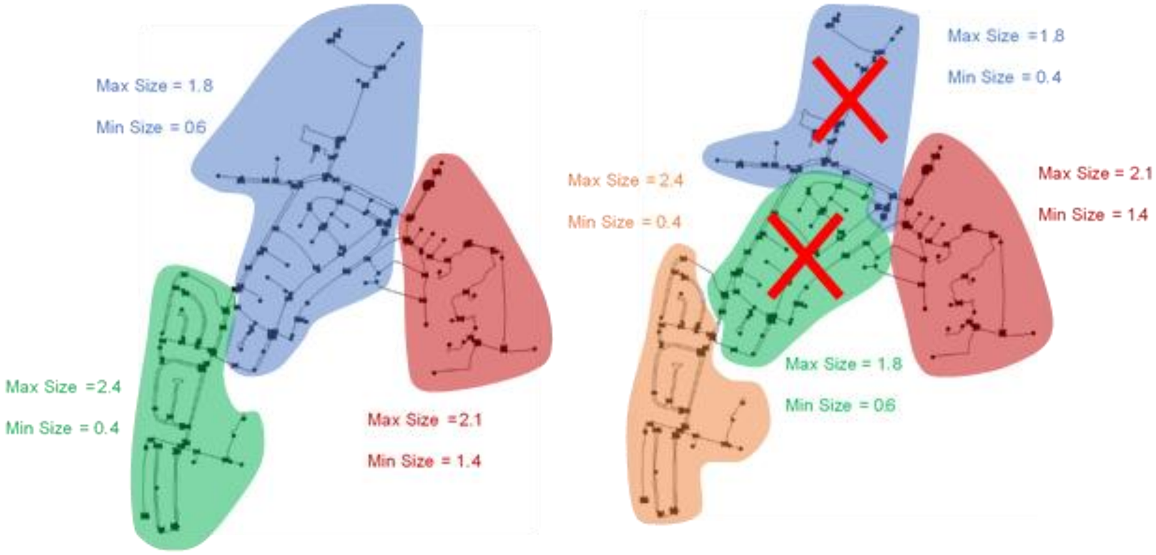
4.4.2.3.8. Assembling the Grouped Sensitivity Matrix

Once the best grouping of leak/burst events has been determined for each sensitivity matrix the final set of leak/burst events which will be considered by the sensor placement technique must be determined. To do this, the groups are checked for each of the final leak/burst events sizes to determine which of the groups should be considered by the sensor placement technique. The first step is to determine the range of valid leak/burst event sizes for the group by considering the maximum and minimum leak/burst event sizes of the individual leak/burst event locations in the group. The maximum and minimum leak/burst event sizes for a group are determined as the largest of the maximum leak/burst event sizes and the smallest of the minimum leak/burst event sizes of the group members. Then, the range of valid leak/burst event sizes for each group are compared to the final leak/burst event size which was used to determine the groups. If the final leak/burst event size falls in the range of a group then that group is valid and should be considered by the sensor placement technique.

The process of checking the validity of groups of leak/burst events is demonstrated in Figure 4-15.



Leak/Burst Event Size = 0.5 Leak/Burst Event Size = 1.0



Leak/Burst Event Size = 1.5 Leak/Burst Event Size = 2.0

Figure 4-16 *Example leak/burst event groups showing which groups to consider for each leak/burst event size*

In Figure 4-15, four different leak/burst event sizes and the groups determined by the leak/burst event grouping technique for each are shown. All groups are valid aside from two groups, shown with a red cross over them, for the leak/burst event size of

2.0. These two groups were not valid because the leak/burst event size was outside of the range of minimum and maximum sizes which were determined for each. From each valid group one leak/burst event location is selected at random and the relevant vector of sensitivities is extracted and entered into a new sensitivity matrix, which is referred to as the grouped sensitivity matrix. The grouped sensitivity matrix is then used by the sensor placement technique to determine the optimal sensor configurations. This sensitivity matrix contains leak/burst events of different sizes and the sizes have been automatically determined considering the accuracy of the instruments and limiting the maximum size of events so that they are not readily apparent. In addition to this the number of duplicate events, considering the sensitivities for each, is reduced so that the minimum number of events is considered.

4.4.2.4. Determining the Optimal Sensor Configurations

4.4.2.4.1. Overview

The optimal sensor placement technique presented in this section is the other main component of the framework of methods. The optimal sensor placement technique is used to automatically determine the best location of pressure sensors in a DMA to maximise the leak/burst localisation performance whilst simultaneously using the minimum number of pressure sensors. There is a trade-off between the number of sensors deployed in a DMA and the level of localisation performance which can be achieved and understanding this trade-off can allow water company personnel to make better decisions about how many sensors to deploy in a DMA. This is related to the ELL (Pearson and Trow, 2005) which, in effect, states that only strategies which can reduce the total cost of detecting, locating and repairing leak/burst events should be employed by water companies so as to provide the best value to customers. In the context of leak/burst localisation this means finding the strategy which can maximise localisation performance whilst minimising the cost of implementing the strategy.

The sensor placement problem is formulated considering two objectives, namely the number of sensors and level of localisation performance which can be achieved using that number of sensors. The critical point of the optimal sensor placement technique is that the leak/burst localisation technique, which was described in section 4.4.1 is used by the optimal sensor placement technique to determine the level of localisation performance for a sensor configuration. This is important as it means that when the leak/burst localisation technique is used to localise real leak/burst events the best performance will be achieved because the leak/burst localisation technique was considered when the sensor configurations were determined. This is not the case when a sensor placement technique is used to deploy sensors but a different leak/burst localisation technique is used to analyse the pressure data to localise leak/burst events.

The problem of determining the optimal sensor configuration(s) is a large and complex one. This is because for typical numbers of potential sensor locations the number of possible combinations of sensors which could be deployed is often extremely large. For a given number of possible sensor locations, s_p, and a desired number of sensors to be deployed, s_d, the number of potential sensor configurations, C, which could be deployed is calculated using Equation 4-5. Equation 4-5 gives the formula for selecting s_d sensors from s_p potential sensor location.

$$C = \binom{s_p}{s_d} = \frac{s_p(s_p - 1)(s_p - 2) \ldots \ldots (s_p - s_d + 1)}{s_d!} \qquad \textbf{(\textit{Equation 4-5})}$$

As an example, deploying 5 sensors in a DMA with 100 potential sensor locations would yield 7 x 10^7 potential configurations of pressure sensors which could be deployed. This example is extremely conservative and, in reality, the number of potential sensor configuration is likely to be many orders of magnitude greater than this. Even for the most computationally efficient sensor placement algorithm is it infeasible that all potential sensor configurations could be searched and evaluated to determine the best ones. Therefore, a strategy for efficiently searching through the potential sensor configurations is required to ensure that the optimal sensor configurations can be determined in a reasonable amount of time.

The solution to this problem which is used by this optimal sensor placement technique is to employ a GA. GAs abound within the field of sensor placement for WDSs because they are able to able to determine optimal (or near optimal) solutions to problems whilst only considering a small fraction of the potential number of solutions, as was covered in section 2.4 of the literature review. In relation to sensor placement this means that a small fraction of the potential number of sensor configurations would need to be considered. The main difficulty is developing suitable objective functions which capture the important performance considerations so that the sensor configurations determined by the optimal sensor placement technique also perform well on real leak/burst events. The remainder of this section will outline the main steps in determining the optimal sensor configurations for a DMA.

4.4.2.4.2. Introduction to Genetic Algorithms

GAs are a class of computational techniques which are inspired by the process of evolution. A summary of some applications of GAs was given in section 2.4 and a comprehensive review was completed by Nicklow et al. (2010). They were developed for the purpose of quickly and efficiently finding solutions to complex problems by iteratively generating, evaluating, selecting and modifying a population of solutions in an attempt to generate better solutions. This process is repeated iteratively so that the solutions to the problem incrementally improve until a satisfactory level of performance is achieved or until some pre-defined parameter, related to the number of solutions to generate and evaluate, is fulfilled. A key characteristic of GAs is that they are a heuristic technique and their performance is not guaranteed for any specific problem. A well-designed and specified GA will generally be able to generate some good solutions for a given problem but this often requires some fine tuning of the parameters of the GA using a trial and error approach. In Figure 4-16, two common parameters, namely the number of solutions to evaluate and the size of the population, are shown but often there can be many other parameters which need to be specified *a priori*. There are several steps to running a GA which are shown in Figure 4-16.

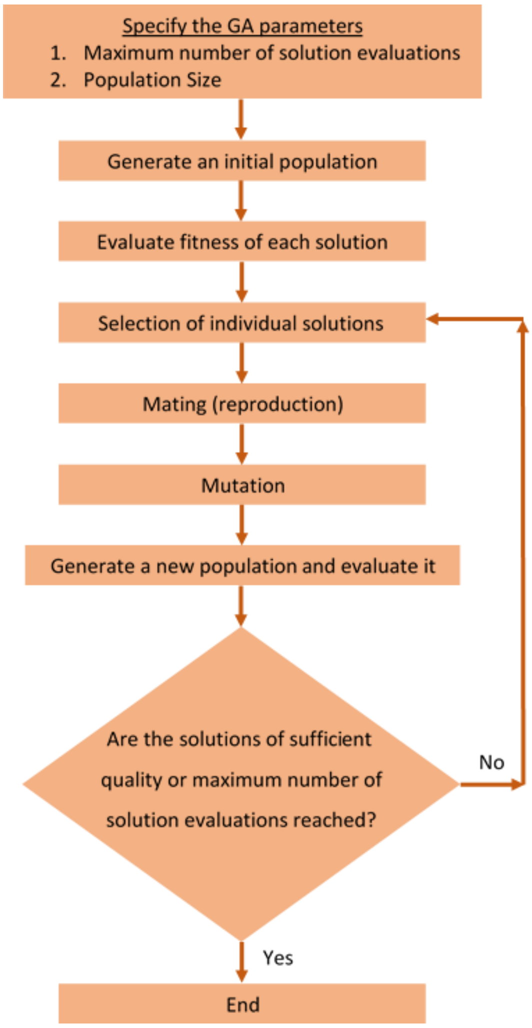


Figure 4-17 *Overview of the process of running a genetic algorithm (adapted from Nicklow et al., 2010)*

The size of the population specifies the total number of solutions which are being considered at each iteration and the maximum number of solution evaluations limits the number of iterations (also called generations) which are performed. The number of solutions is an important factor in determining the amount of time that a GA will need to complete a run and is often selected to minimise the total time taken whilst ensuring that solutions of sufficient quality are generated. Each member of the population is comprised of a set of decision variables. The decision variables are the quantities to be determined by the GA to ensure that adequate performance is achieved. A relevant example for WDSs would be to vary the pipe sizes in a hydraulic model to minimise the cost of designing a new, or expanding an existing, WDS to satisfy some performance constraint such as providing adequate pressure (Savic and Walters, 1997). In this case the size of each individual pipe is one decision variable and the best combination of pipe sizes would give the minimum cost. The number of decision variables and the allowable values for each decision variable is governed by the problem to be solved.

The first step in running a GA is to generate an initial population. There are several approaches which can be used including randomly selecting the value of each decision variable, specifying the values of each decision variable manually or using a combination of the two. Once the initial population has been set the GA then evaluates each of the members of the population. To do this, one or more objective functions (or fitness functions) can be used. The objective functions allow the quality of each population member to be determined and are designed considering the required outcome of the GA run. A common bi-objective function will aim to minimise cost whilst simultaneously maximising a different measure of performance where the two objectives are competing. After the initial population has been evaluated against the objective(s) then the best solutions are selected and will be used to generate the next population. To generate a new population the values of the decision variables of the selected population members are modified. There are a wide variety of methods for modifying the decision variables and the combination of operators to use is a key decision to make when designing a new GA. To ensure that the modification of the existing best solutions does not degrade the performance some unmodified solutions are usually included in the next population. Whichever method is used to generate the new population, all of its members are evaluated against the objective functions again. This process is repeated until either satisfactory performance is achieved against the objectives or the pre-specified number of solutions have been generated and evaluated. The next section will introduce the GALAXY MOEA which is used to determine the optimal sensor configurations by the optimal sensor placement technique.

4.4.2.4.3. GALAXY Multi Objective Evolutionary Algorithm for Determining the Optimal Sensor Configurations

The GALAXY MOEA was developed by Wang et al. (2017) for solving WDS optimisation problems. This section will outline the main features of the GALAXY MOEA. For more in depth information regarding the operation of the GALAXY MOEA and its application to several well-known benchmark problems pertaining to WDSs please refer to Wang et al. (2017).

There are many GAs which have been applied to multi-objective problems including NSGA-II (Deb et al., 2002), which is the most widely used GA applied to WDSs. The main drawback of most GAs is the need to specify and fine-tune several parameters to ensure that the performance of the MOEA is acceptable for the problem being considered. Given that the solution space can vary drastically for different problems this means that significant time and effort are spent in determining the best values of the GA parameters. The GALAXY MOEA, however, only requires two parameters need to be specified. The two parameters that must be specified prior to running the GALAXY MOEA are the size of the population and the number of functions evaluations (NFEs). The NFEs is the number of individual population members which are evaluated using the objective functions and is equal to the population size multiplied by the number of generations. The fact that only two parameters need to be specified alleviates the need to systematically determine the best values of the parameters. This is particularly beneficial because the best parameter values could potentially need to be determined for every different DMA which would represent a significant time investment.

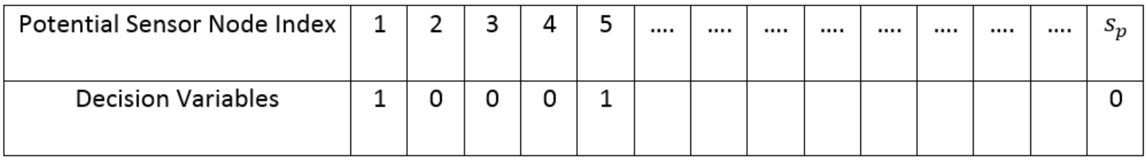
Aside from overcoming the parameterisation issue there are several other noteworthy features of the GALAXY MOEA which make it a good candidate for optimal sensor placement in WDSs. Firstly, the GALAXY MOEA is a hybrid evolutionary algorithm that it makes use of search operators from several different types of evolutionary algorithms during the solution generation stage. These search operators and the algorithms from which they were taken are; turbulence factor (AMALGAM – particle swarm optimisation), differential evolution, simulated binary crossover for integers and uniform mutation (GA - NSGA-II), Gaussian mutation and dither creeping. The combination of search operators governs how new solutions are generated for each new population and affects the ability of an evolutionary algorithm to efficiently cover the search space to identify good solutions in a reasonable timeframe. Wang et al. (2017) stated that using multiple search operators can allow search operators with different performance characteristics to drive the search for the global optimal solutions at different stages and that this may aid in preventing the GALAXY MOEA from getting stuck on a local optimal solution. Additionally, some of the search operators are better at allowing the solution space to be covered whereas others are better at local searches.

The GALAXY MOEA makes use of a hybrid strategy to determine which members of the intermediate population (comprised of some members of the current population and the offspring generated by modifying members of the current population) to include in subsequent populations. The ε-replacement or the fast nondominated sorting strategy used by NSGA-II (Deb at al., 2002) are used depending upon how many nondominated solutions are contained in the intermediate population. On top of this the GALAXY MOEA uses the genetically adaptive strategy which was used by AMALGAM to favour the best performing search operators by allowing more solutions generated using these search operators into the next population (Wang et al., 2017) and not considering as many new solutions generated by the lesser performing search operators.

The GALAXY MOEA was applied to five benchmark optimisation problems of various sizes and the performance using four metrics was compared to three well-known evolutionary algorithms, namely, BORG (Hadka and Reed, 2015), NSGA-II (Deb et al., 2002) and ε-MOEA (Deb et al., 2005). The four performance metrics used to evaluate all of the evolutionary algorithms were concerned with convergence to the best-known Pareto fronts for each problem, diversity of the population to ensure good coverage along the entire width of the known Pareto front and that consistent convergence was achieved across the entire width of the Pareto front. The key finding presented by Wang et al. (2017) is that the GALAXY MOEA was able to perform as well or better than the three other evolutionary algorithms given the same computational budget and considering 30 runs of each MOEA for each of the five problems considered. This important finding was independently confirmed by Wang et al. (2020).

4.4.2.4.4. Sensor Placement Decision Variables

In the context of the optimal sensor placement technique the decision variables are the locations of the sensors. The GALAXY MOEA decision variables can only take integer values and each decision variable must have the same range of possible values. Therefore, the sensor configuration is denoted by a binary vector of ones and zeros. The length of the binary vector of sensor locations is equal to the number of possible sensor locations. For each decision variable a one at the corresponding index of the sensor location is considered to indicate its membership of part of the sensor configuration being considered for that member of the population. This is illustrated in Figure 4-17 below.



Potential Sensor Node Index	1	2	3	4	5	....	....	....	....	....	....	....	....	s_p
Decision Variables	1	0	0	0	1									0

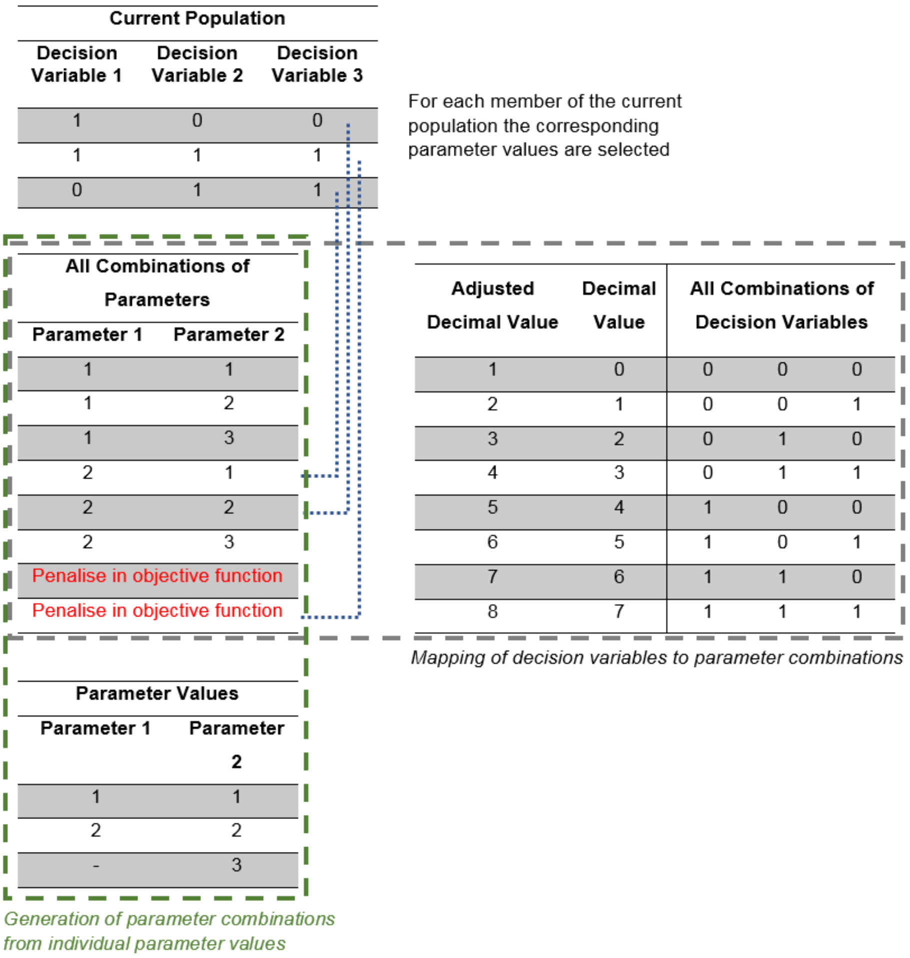
Figure 4-18 *Example sensor placement decision variables*

4.4.2.4.5. GALAXY Parameter Selection

A simple process for selecting the GALAXY MOEA parameters has been used by the optimal sensor placement technique. The case studies provided in Wang et al. (2017) gave some indicative values for both the population size and the NFEs based upon the number of decision variables and the size of the solution space. The size of the solution space is equal to the total number of possible solutions which could be generated considering the number of decision variables and the number of values that each decision variable can take. The size of the solution space is equal to the number of values each decision variable can take raised to the power of the number of decision variables. The GALAXY MOEA parameters primarily affect the speed of determining the optimal (or near optimal) solutions and the coverage of the Pareto front. In the context of the sensor placement objective function which uses the number of sensors (a discrete variable) as one of the objectives, covering the entire width of the Pareto front along this objective is simple because there are relatively few permissible values that it can take due to this being constrained. The selection of both parameter values, therefore, is aimed at ensuring that the GALAXY MOEA will produce acceptable results in a reasonable computational time. Measuring convergence in this case is problematic as to do this a known Pareto front is required against which convergence can be measured. Wang et al. (2017) dealt with this by measuring the amount of change between successive generations rather than how well converged towards the best-known Pareto front the latest population was. In the same way, the sensitivity analysis for determining suitable values for both GALAXY parameters measured the amount of improvement which was achieved by successive generations of the GALAXY MOEA. When the improvement in performance was diminished then the GALAXY MOEA was assumed to have converged. A value of 0.1%, calculated by summing the localisation performance of all non-dominated solutions in the latest two populations was used. The number of functions evaluations was then selected by running the sensor placement technique using several DMAs to ensure that less than 0.1% performance improvement was attained in those cases. The values of 100,000 was then selected for use with the sensor placement technique. The population size affects the rate of convergence because choosing a larger population size will normally increase the diversity in the population which allows better coverage of the solution space. For the sensor placement technique, a number of different population sizes were tested and a population size of 200 was found to be suitable. The GALAXY parameters selected were not necessarily the ideal values but were found to lead to acceptable performance levels for the sensor placement technique as a whole.

4.4.2.4.6. Automatic Leak/Burst Localisation Parameter Selection

Aside from the parameters for the GALAXY MOEA the two parameters used by the leak/burst localisation technique also need to be determined. The leak/burst localisation parameters are used by the sensor placement objective function as part of the leak/burst localisation technique. The leak/burst localisation parameters must be selected prior to running the leak/burst localisation technique. The problem, however, is that the performance of each sensor configuration depends upon the values of both leak/burst localisation parameters. In addition to this, the best value of the parameters can be different for each sensor configuration. To address this problem the leak/burst localisation parameters are considered as part of the sensor placement problem whose best values are to be determined at the same time as the sensor configurations. Decision variables considered by the GALAXY MOEA are constrained to take only integer values with the same range for all decision variables. Therefore the decision variables must be mapped to the possible parameter combinations as show in Figure 4-18.



Current Population		
Decision Variable 1	**Decision Variable 2**	**Decision Variable 3**
1	0	0
1	1	1
0	1	1

For each member of the current population the corresponding parameter values are selected

All Combinations of Parameters	
Parameter 1	**Parameter 2**
1	1
1	2
1	3
2	1
2	2
2	3
Penalise in objective function	
Penalise in objective function	

Adjusted Decimal Value	**Decimal Value**	**All Combinations of Decision Variables**		
1	0	0	0	0
2	1	0	0	1
3	2	0	1	0
4	3	0	1	1
5	4	1	0	0
6	5	1	0	1
7	6	1	1	0
8	7	1	1	1

Mapping of decision variables to parameter combinations

Parameter Values	
Parameter 1	**Parameter 2**
1	1
2	2
-	3

Generation of parameter combinations from individual parameter values

Figure 4-19 *Generating the combinations of parameters and selecting them using binary decision variables using the GALAXY MOEA*

The decision variables pertaining to the sensor locations take a value of zero or one therefore the decision variables to select the parameters must be constrained in the same way. The interpolation exponent can take integer values greater than zero. The search area threshold is a real number between zero and one so some modification of the decision variables is required to allow simultaneous selection of the sensor locations and the leak/burst localisation parameters. The proposed solution is to generate a table of all allowable combinations of the leak/burst localisation parameters and then use a combination of ones and zeros to generate a binary number. The binary number will be used to select the combination of parameters as a row from the table of parameter combinations. Invalid row indices will be penalised to ensure that only the values related to valid parameter combinations are permitted as shown in Figure 4-18. The number of additional decision variables must be determined prior to running the sensor placement technique. This is because the number of decision variables required depends upon the number of parameter combinations which are being considered by the sensor placement technique. The range of allowable values for the interpolation exponent was chosen to be significantly wider than the range of values found in the literature. The interpolation exponent can take any integer value between 1 and 30, inclusive. For the search area threshold, the allowable values were between 0.5 and 0.95 in 0.05 increments, giving 10 allowable values. This ensures that only junctions with an interpolated value above the mean of the interpolation surface will be included in the search area. Given the allowable values for both of the leak/burst localisation parameters which are being used there are 300 parameter combinations. To accommodate all of the parameter combinations the number of additional decision variables (n_{adv}) can be determined according to Equation 4-6.

$$n_{adv} = \lceil \log_{n_{vdv}} C_{dv} \rceil \qquad \textbf{\textit{(Equation 4-6)}}$$

Equation 4-6 states that the number of additional decision variables is the ceiling (rounding up to the nearest integer) of the logarithm of the number of combinations of parameters (C_{dv}) using a base equal to the number of values each decision variable can take (n_{vdv}). Each decision variable can take two values (either a one of a zero). To accommodate 300 combinations of parameters 9 additional decision variables are needed. Using 9 additional decisions variables gives a total of 512 possible combinations of decision variables so those combinations pertaining to the numbers 301-512 are penalised so that they are removed from future populations and only solutions which relate to the rows of the table of parameter combinations are included in subsequent populations. The vector of decision variables is extended so that in addition to the sensor locations the decision variables for the parameters are also considered simultaneously. The inclusion of the decision variables was tested against the best-known combinations of parameters to ensure that the same parameter combinations were identified using the automatic procedure outlined here as for manually evaluating all parameter combinations using a single fixed combination for each sensor placement run. The results indicated that the best combination of parameters was determined for across the entire Pareto front. The results of this comparison are included in case study 5.6 in chapter 5.

4.4.2.4.7. Performance Metrics

One problem which has been identified is that there is a lack of consistency between how the performance of different leak/burst localisation/sensor placement techniques are reported. This is understandable as there can be great variation between the techniques but it means that comparing the performance of two leak/burst location/sensor placement techniques is difficult. Several of the performance measures which are used by leak/burst localisation/sensor placement techniques are given below:

- Number of leak/burst events correctly localised

An important consideration is how many leak/burst events are correctly localised for a leak/burst localisation technique or a sensor placement technique which is used for leak/burst localisation. A common approach is to identify an area or a set of leak/burst event locations which is suspected of containing the leak/burst event. Then when the actual leak/burst event location is confirmed then if it fell within the area which was suspected it is considered to have been correctly localised. The number (or percentage) of leak/burst events which were correctly localised is used to determine how good a leak/burst localisation technique or sensor configuration is. The goal is to successfully localise as many leak/burst events as possible. This is the most commonly used metric for evaluating the localisation performance of leak/burst event localisation and sensor placement techniques.

- Localisation accuracy (size of search area)

Localising many leak/burst events does not necessarily ensure that the leak/burst localisation technique or sensor configuration will actually provide a meaningful benefit to a water company. This is because by highlighting a large area it becomes easier to successfully localise a leak/burst event. However, by highlighting a larger area the potential benefit to a water company is significantly diminished. One benefit of localisation is that it can guide field teams closer to the location of a leak/burst event and reduce the amount of time taken to find a leak/burst event. Reducing the time to find leak/burst events means that in a given time period more leak/burst events can be repaired which is more efficient and reduces the cost of repairing leak/burst events. Therefore, guiding field teams to a small area containing the leak/burst event ensures that they will find it quickly which is most beneficial. There is a trade-off between the number of leak/burst events which are correctly localised and the size of the area which is highlighted for a leak/burst event. Surprisingly, the localisation accuracy has only been used by a small number of techniques.

- Distance to the leak/burst event location

A similar concept to the size of the search area is the distance between the suspected leak/burst event location and the actual leak/burst location. The further apart the two are the worse the localisation performance is deemed to be. This method is used where only a single leak/burst event location is suspected by a given leak/burst localisation technique. Generally, the distance travelled along the pipes is considered as it is more meaningful than the Euclidean distance. The techniques which use the distance tend to be based upon model-based fault diagnosis.

- Number of distinct search areas

Another consideration is that for many leak/burst localisation techniques multiple different areas in a DMA can be highlighted. This is true for leak/burst localisation techniques which make use of the leak/burst signature or the sensitivity to infer the location of the leak/burst event. Highlighting multiple areas is problematic for two reasons. The first is that it undermines confidence in the results which are provided to field teams, particularly if the distance between the two search areas is great. Secondly, it incurs a cost of travelling between the areas if both need to be searched, which adds to the time taken to find the leak/burst event. It is therefore desirable that, for as many leak/burst events as possible, that only a single search area is highlighted to prevent these problems from occurring. Only a single technique made use of the number of search areas and also accounted for the distance between the search areas to evaluate localisation performance.

- Resilience to sensor faults/failures

Another criterion which arises as a result of the fact that sensors and communications technologies can incur faults or failures, particularly as they age, is how well a leak/burst event localisation technique or sensor placement technique copes with these faults/failures. Even in the face of increased asset management by water companies in recent times sensor faults and failures will inevitably occur. Some techniques account for this but it is by no means ubiquitous.

Of the performance metrics described here the most important are the number of events correctly localised and the accuracy. The distance between the actual leak/burst event location and the suspected leak/bust event location is deemed less appropriate than the accuracy because it does not work when multiple leak/burst event locations are included in the search area. Finally, given that an interpolation technique is being used by the leak/burst localisation technique there is some inherent resistance to sensor failures built into the localisation technique when compared to those techniques using the leak signature as part of a model-based fault diagnosis leak/burst localisation technique. There are other means, for example a rigorous maintenance regime, which can help to alleviate the problem of sensor faults and failures. In the coming sections, the two objective functions which are used by the optimal sensor placement technique are outlined. One of the objectives is related to the cost of a sensor configuration and the other is related to the localisation performance.

4.4.2.4.8. Objective Function 1 – Localisation Performance

The first objective function used by the optimal sensor placement technique evaluates the ability of a sensor configuration to localise all possible leak/burst event scenarios which are contained in the grouped sensitivity matrix which is produced by the leak/burst event grouping technique. It combines several of the performance metrics outlined in section 4.4.2.4.7 to determine an average localisation performance considering all of the leak/burst event scenarios which are contained in the grouped sensitivity matrix. For each leak/burst event the size of the search area produced by the leak/burst localisation technique is determined. Once the leak/burst localisation technique has been performed a group of graph nodes (which represent junctions in the DMA) is highlighted based upon the values which have been estimated for all junctions in the graph for the DMA. From this group of graph nodes the relevant group of graph links (which represent the pipes in the DMA) contained in the search area can be determined by considering any link which has both of its end nodes contained in the search area as being in the search area. The total weights of all graph links, which is the same as the length of the pipes that they represent, in the search area is allocated as the size of the search area. The reason for using the pipe length as opposed to the number of nodes is that it gives a better representation of the amount of the DMA which would need to be searched. This is because the density of junctions in the hydraulic model, which is used to create the graph along which the leak/burst localisation technique is performed, can be highly variable. If a sensor is placed in an area with a high density of junctions then many more junctions are likely to be in the search area which artificially increases the size of the search area. This would dissuade the sensor placement technique from placing sensors in areas where there are many junctions even though the sensor location could actually be good for leak/burst localisation. By considering the length of the links instead, this problem is overcome.

The next step is to apply a penalty to the size of the search area for a leak/burst event which is not correctly localised. This means that the modelled leak/burst location is not one of the nodes in the search area produced by the leak/burst localisation technique. The purpose of this penalty is to steer the population generated by the GALAXY MOEA towards sensor configurations which correctly localise the greatest number of leak/burst events. The penalty replaces the size of the search area given by the leak/burst localisation technique with the total length of links in the graph to mimic the fact that if the leak/burst is not correctly localised that the entire DMA may need to be investigated by the field team prior to finding it. The penalty for incorrect localisation is demonstrated in Figure 4-19. In this case the size of the red area is added to the size of blue area to give a total penalised value of the objective function.

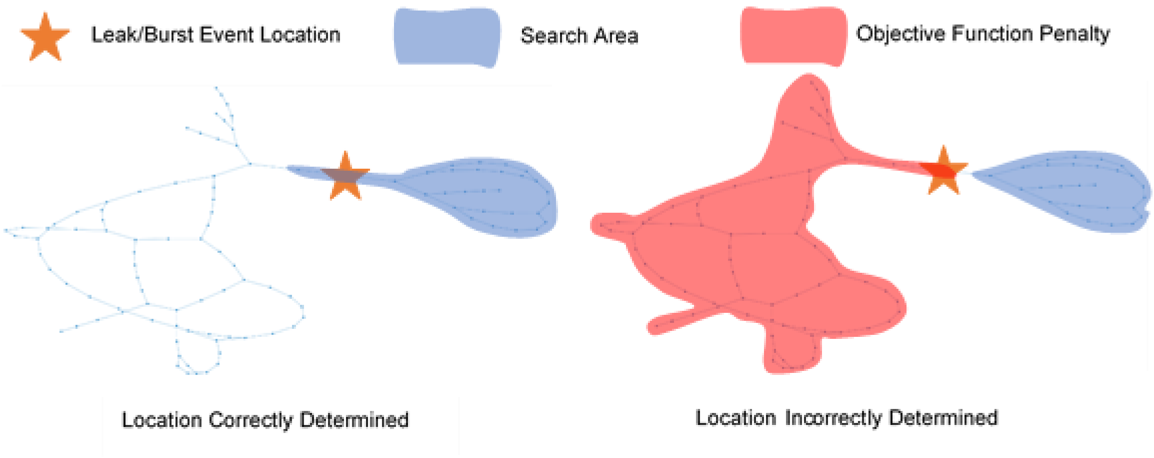


Figure 4-20 *Penalty for incorrect localisation of a leak/burst event used by the optimal sensor placement technique*

A second penalty is also applied for any leak/burst event for which multiple search areas are produced by the leak/burst localisation technique. To apply this penalty a graph-based procedure, which uses the same graph as the leak/burst localisation technique has been developed. The penalty for multiple search areas is demonstrated in Figure 4-20. The size of red area is added to the size of the blue area as for the first penalty. If more than one search area is produced for a leak/burst event then, in the same way as for incorrect localisation, all of the graph links are counted as being in the search area. Both penalties can be applied to the same leak/burst event meaning but in this case the same value for the size of the search area is determined as for the cases when either of the penalties is applied individually. After all of the leak/burst events have been analysed and the size of the search area for each (including any penalties) has been determined for each leak/burst event then the mean size of the search areas is calculated. Finally, the mean size of the search area is normalised, using the total length of the pipes in the DMA graph to give a value between zero and one which is used to measure the localisation performance of each sensor configuration.

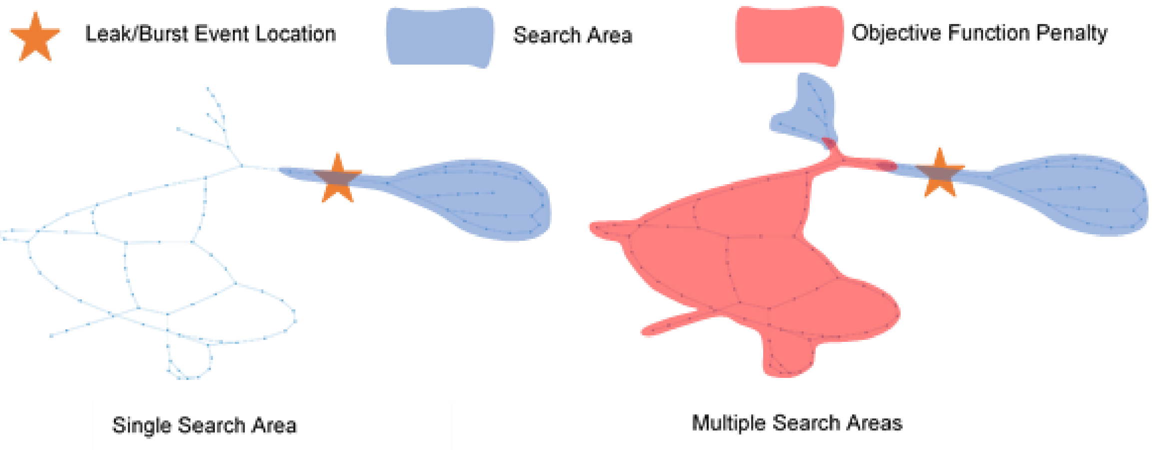


Figure 4-21 *Penalty for determining multiple search areas of a leak/burst event used by the optimal sensor placement technique*

To count the number of search areas a list of the graph nodes, and hence the graph links, which are in the search area is extracted. From this, a new graph is created and the number of disconnected subgraphs in the new graph can be determined using the function "conncomp" in Matlab. This process is demonstrated in Figure 4-21.

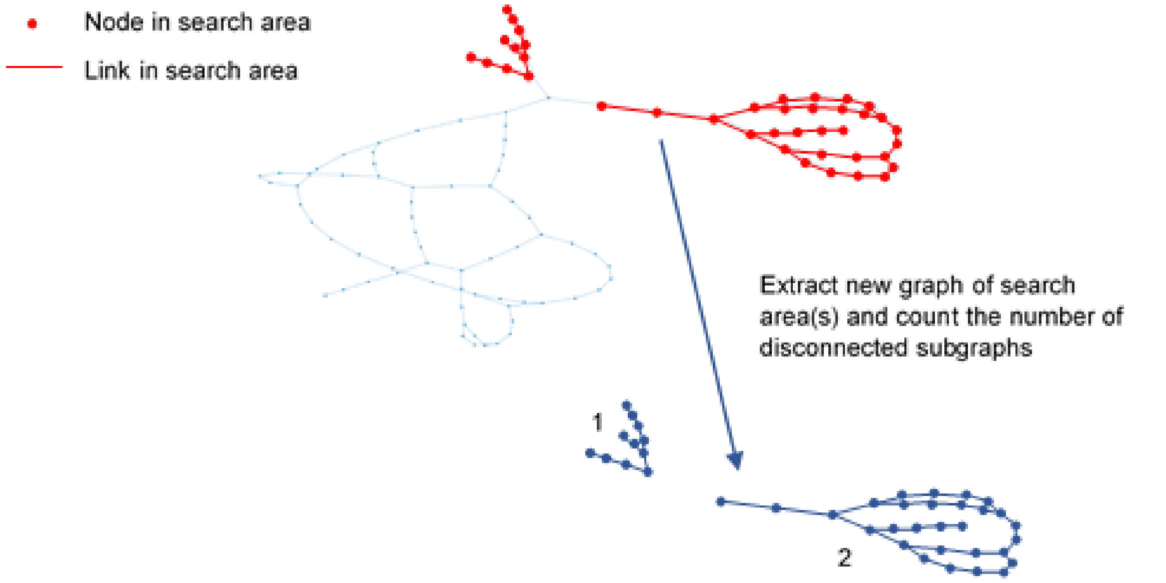


Figure 4-22 *Procedure for determining when multiple search areas are determined for a leak/burst event*

4.4.2.4.9. Objective Function 2 – Cost (Number of Sensors)

The second objective function used by the sensor placement technique is related to the cost of each sensor configuration generated by the GALAXY MOEA. There are many factors which dictate the cost that will be incurred by installing sensors in a WDS which were outlined in section 4.4.2.3.4. It has been assumed by the sensor placement technique that the cost is solely dependent upon the number of sensors which are being installed. Furthermore, a linear relationship between the number of sensors and the cost, assuming no economies of scale, has been used. To simplify this the number of sensors is used directly as the second objective. In reality, there are other factors which also contribute to the overall cost, which are independent of the number of sensors. One example is the data communication and storage which may require a base level of investment which is not strictly proportional to the number of sensors. Accounting for all of the costs would be extremely time consuming as the total cost is dependent upon many variables which cannot easily be estimated.

4.4.2.4.10. Sensor Placement Constraints

Due to the potential size of the sensor placement problem for a typical DMA it is necessary to implement some constraints to reduce the size of the problem and focus the search on the most practically feasible solutions.

Number of Sensors

The first constraint applied to the sensor placement problem is to select the maximum and minimum number of sensors which can be installed in a DMA. There is a theoretical maximum number of sensors which can practically be installed due to trade-off between the costs and benefits of implementing the framework of methods. The sensor placement technique determines the shape of the Pareto front (the curve of trade-off between multiple objectives) to allow water company personnel to select the appropriate number of sensors to deploy in a DMA. To understand where the point of diminishing returns, where adding another sensor does not provide great enough benefit to justify its addition, sits on the Pareto front between the two objectives it is necessary to allow the sensor placement technique to consider higher numbers of sensors that would reasonably be installed in a DMA. The reason for limiting the maximum number of sensors is to reduce the size of the solution space which is possible. Allowing smaller numbers of sensors to be installed significantly reduces the size of the solution space and ensures that only feasible solutions, in terms of the number of sensors deployed, are considered. For all sensor placement runs which have been considered a maximum value of ten sensors has been used.

A constraint is also placed on the minimum number of sensors which should be considered and there are two reasons for this. The first reason is that using too few sensors results in extremely poor performance and considering these solutions diverts time and computational effort away from focussing on solutions which produce an acceptable level of performance. It is well known within the fields of spatial analysis and geostatistical techniques that using too few measurements will not produce meaningful results and the leak/burst localisation technique is no different in this respect. The minimum number of sensors to be considered by the sensor placement technique has been set to three sensors although this number can be increased or decreased depending upon the requirements of the user.

Leak/Burst Localisation Parameters

The allowable range of values which are considered for each of the leak/burst localisation parameters are specified prior to running the sensor placement technique. The total number of parameter combinations is a product of the number of allowable values for the interpolation exponent and the search area threshold. Selecting the range of allowable values for each leak/burst localisation parameter considers two important factors. The first factor is that allowing a larger range of allowable values for each will increase the maximum possible performance that can be achieved by the sensor placement technique. This is because allowing too few values will increase the likelihood that the best theoretical values, which cannot be known *a priori,* are not considered by the sensor placement technique. However, given that the leak/burst localisation parameters are considered as decision variables by the sensor placement technique, allowing more values for both parameters increases the size of the search space by a factor proportional to the square of the increase in the number of allowable values. A sensitivity analysis was conducted to investigate the best combination of leak/burst localisation parameters using two simple hydraulic models. A summary of this sensitivity analysis is provided in case study 5.5 in Chapter 5. The allowable values are given in Table 4-3. The allowable values must be between the maximum and minimum allowable values and a multiple of the increment.

Table 4-3 *Allowable values for the leak/burst localisation parameters*

Parameter	**Minimum Allowable Value**	**Maximum Allowable Value**	**Increment**
Search Area Threshold	0.5	0.95	0.05
Interpolation Exponent	1	30	1

4.5. Summary

In this chapter a novel framework of methods has been described which addresses the problem of localising new leak/burst events by determining optimal configurations of pressure sensors in a DMA and analysing the data collected from them to infer the approximate location of new leak/burst events.

The key benefits of the framework of methods are summarised in the bullet points below:

- The integration of the leak/burst localisation technique and the sensor placement technique ensures that the localisation performance which is achieved by the leak/burst localisation technique is maximised by minimising the size of the area determined for a new leak/burst event. The optimal sensor configurations determined by the sensor placement technique are therefore optimal with respect to the leak/burst localisation technique.

- The leak/burst localisation technique features a newly developed SC-IDW interpolation technique which considers the length and connectivity between pipes in the DMA to approximate the leak/burst event location. For every junction in the DMA graph the change in pressure for multiple sensors in used to estimate the change in pressure. This means that the leak/burst localisation technique is more robust to sensor failures than the model-based fault diagnosis techniques described in the literature such as Farley et al. (2010) or Perez et al. (2011).

- The leak/burst localisation technique identifies a sub-region of the DMA as opposed to an individual leak/burst event location which ensures it is robust to the uncertainty which is contained in the pressures in a real DMA. Calculating the changes in pressure during the night and using multiple days to determine the average normal behaviour of the pressure in the DMA also ensures robustness to this uncertainty.

- No comparison between the hydraulic model data and the data collected from a real DMA is used to infer the approximate leak/burst event location. This reduces the reliance on a highly calibrated hydraulic model which is used by many leak/burst localisation techniques. Even leak/burst localisation techniques which did not use a highly calibrated model (Farley et al., 2010) still require a comparison reference to the idealised search areas produced by a hydraulic model in order to infer the approximate location of a new leak/burst event in a real DMA. This is overcome by using the leak/burst localisation technique presented in this chapter.

- The sensor placement technique directly considers the leak/burst localisation accuracy which is penalised to ensure that the maximum number of leak/burst events are localised correctly whilst minimising the average size of the search areas produced using the leak/burst localisation technique. Sensor configurations which identify multiple search areas are heavily penalised to steer the sensor placement technique towards sensor configurations which reduce the occurrence of this problem. This improves on previous techniques which only consider the number of leak/burst events correctly localised such as Casillas et al. (2013, 2015).

- Using a bi-objective formulation of the sensor placement problem, which simultaneously minimises the number of sensors required whilst maximising leak/burst localisation performance, ensures that the sensor configurations which provide the maximum cost-benefit are identified. A Pareto front of optimal sensor configurations are determined, in a single run of the sensor placement technique, from which the preferred solution can be selected.

- A novel procedure for automatically determining the range of leak/burst event sizes which should be considered for each DMA was developed. This uses the accuracy of the pressure instruments being used to determine the smallest leak/burst event sizes and limits the maximum event size by specifying a maximum allowable increase in flow relative to the daily peak. This removes the need for a detailed investigation prior the running the sensor placement technique and allows multiple leak/burst event sizes to be considered by the sensor placement technique.

- A novel procedure for grouping leak/burst events was developed to reduce the number of duplicate leak/burst events considering the changes in pressure for all candidate sensor locations only and not the leak/burst event locations directly. The accuracy of the instrument is also accounted for to ensure that only events which cannot be differentiated by a complete sensor configuration are grouped together. This means that for multiple leak/burst event sizes that the entire sensitivity matrix (for each event size) does not need to be considered. This significantly reduces the number of leak/burst event scenarios which are considered by the sensor placement technique.

Chapter 5 - Case Studies for Verification of the Framework of Methods

5.1. Introduction

The framework of methods described in Chapter 4 was developed iteratively and several case studies were conducted throughout the development process to verify its performance and suitability at key points. A combination of hydraulic model data and engineered event data was used for verifying the framework of methods. In addition to this, hydraulic models of varying degrees of complexity were used at different points throughout the development of the leak/burst localisation technique. Each case study is now described in detail to outline its purpose within the context of the development of the framework of methods, the analysis procedure which has been used and the results that were achieved. This chapter will detail the case studies used during the development of the framework of methods. The leak/burst localisation technique was developed first because it is used as a central part of the sensor placement technique and the case studies pertaining to the leak/burst localisation technique are thus demonstrated first.

5.2. Overview of Verification Hydraulic Models

To verify the performance of the leak/burst localisation technique three hydraulic models were used. The three hydraulic models, referred to hereafter as the verification hydraulic models, are shown in Figure 5-1. Two smaller hydraulic models were derived from a hydraulic model of a real DMA with some sections removed to reduce their size and complexity. One dendritic and one looped hydraulic model were used to evaluate the performance of the leak/burst localisation technique. This is because different connectivity between the pipes in a DMA will influence the changes in pressure which are seen. The purpose of the two smaller verification hydraulic models was to allow rapid testing of changes made to the leak/burst localisation technique in order to identify beneficial changes at an early stage in the development process. In addition to the two simple hydraulic models a hydraulic model of a real DMA (DMA 123-09) located in North West UK was used for further testing to demonstrate the performance of the developments made to the leak/burst localisation technique on a realistic hydraulic model.

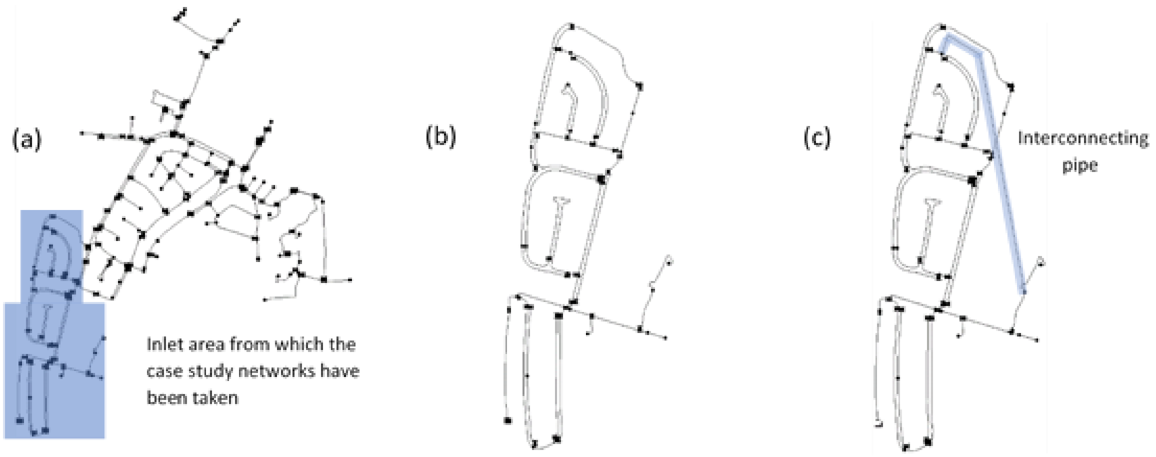


Figure 5-1 *The (a) DMA 12309 verification hydraulic model, (b) “dendritic” verification hydraulic model and (c) “looped” verification hydraulic model with the added interconnecting pipe highlighted in blue. The networks in (b) and (c) are taken from the area in (a) which is highlighted in blue*

5.3. Leak/Burst Localisation Case Studies

The purpose of the case studies in this section is to demonstrate the performance of the leak/burst localisation technique as a whole. Key developments in the leak/burst localisation technique and the resulting performance benefits are included to justify their inclusion in the framework of methods.

5.3.1. Case Study 5.1: Comparison of Spatially Constrained and Euclidean Distance Functions

One of the key developments of the SC-IDW interpolation technique described in section 4.4.1 is the use of a spatially constrained distance function. Traditionally, the distance used by IDW interpolation is the Euclidean distance between two points (Zimmerman et al., 1999). The SC-IDW interpolation uses the distances travelled along the pipes in a DMA to perform the interpolation. Case study 5.1 demonstrates the benefit of using the SC-IDW interpolation technique when compared to using Euclidean distance, as per the traditional variant of IDW interpolation. Presented below are the localisation results for an engineered event which was conducted in a real DMA (DMA 123-09) in United Utilities WDS. DMA 123-09 contains 1664 properties and is fed under gravity from a nearby service reservoir without passing through any other DMAs. The engineered event was conducted on 18/12/2015 at the location depicted by the red triangle. To perform the engineered event a hydrant was opened to achieve an approximate flowrate of 0.5l/s. The flow rate of the engineered event was measured using a flowmeter installed in a standpipe connected to the opened hydrant. The hydrant was opened at approximately 01:30 and closed at 04:00 to coincide with the period of minimum demand in the DMA which is also used by the leak/burst localisation technique to determine the approximate location of leak/burst events. Five pressure sensors were deployed in the DMA prior to the engineered event at the locations depicted by black circles in Figure 5-2. The sensors locations were chosen using engineering judgement rather than using a sensor placement technique.

Three different approaches were used to determine search areas which are shown in Figure 5-2. In the first case, which is labelled (a), the search area was estimated by drawing an equidistant locus (also called a "buffer" in GIS) around pressure sensor 3 which measured the largest deviation in pressure when compared to the normal operating pressures. The normal operating pressures were calculated using the procedure outlined in section 4.4.1.3. Figure 5-2(a) depicts an approximate search area for the engineered event because field teams would only have the pressure and flow deviations for any instruments installed in the DMA.

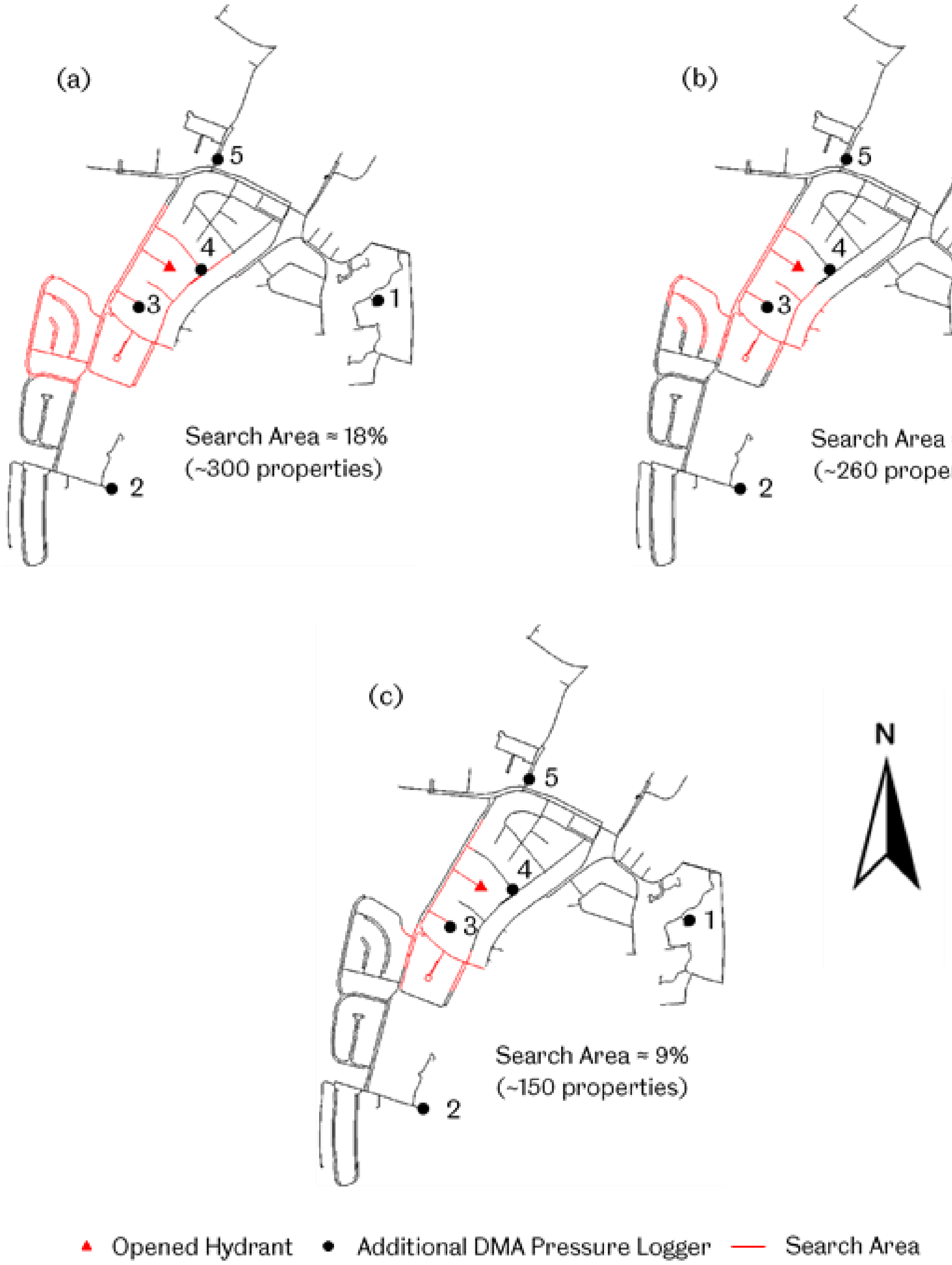


Figure 5-2 *Search area produced by (a) approximating the area to be searched by field teams (b) traditional IDW interpolation (c) SC-IDW interpolation*

In Figure 5-2(b), the search area produced when the leak/burst localisation technique was run using traditional IDW interpolation, with the Euclidean distance function, is depicted. The final case, shown in Figure 5-2(c), is the case where the leak/burst localisation technique was run using the spatially constrained distance function. In all three cases the search area is depicted by the red pipes. The red area in Figure 5-2(c) would be passed to field teams to guide them towards the approximate location of the leak/burst event as part of the leak/burst localisation technique.

From the search areas shown in 5-2(b) and 5-2(c) several interesting points can be observed. Using the spatially constrained distance function rather than the Euclidean distance function led to a reduction of almost 50% in the size of the search area so that only 9% of the DMA would need to be searched rather than 16%. This was largely achieved by removing the area to the Southwest of sensor 3 which was furthest away from the engineered event location. This area was removed because it was close to sensor 3 in the Euclidean sense but significantly further away when the distance travelled along the pipe was considered. Another interesting benefit of using the spatially constrained distance function is that a single contiguous search area is produced rather than several disconnected search areas. There are two clear examples of two disconnected pipe sections included in the search area in Figure 5-2(b). These are highlighted in Figure 5-3 for clarity. Whilst these two sections do not contribute a significant portion of the search area in this case it shows the limitation of using the Euclidean distance function in the leak/burst localisation technique.

Another beneficial trait of the spatially constrained distance function is the removal of some pipe sections between sensors 3 and 4 which are both in close proximity to the engineered event location. The sections of pipe which are closer to sensor 3 according to the distance travelled down the pipes are removed from the search area because the deviation in pressure at sensor 4 was lower than for sensor 3. This focussed the search area more closely around sensor 3 whilst still including the engineered event location in the search area. Related to this is the ability of the SC-IDW interpolation to differentiate between parallel pipes. For the search area shown in Figure 5-2(c) one half of a parallel pipe section was removed from the search area whilst the other half remained in the search area. Removal of this section of pipe alone could have contributed noticeably to the reduction in the time to find the leak/burst event.

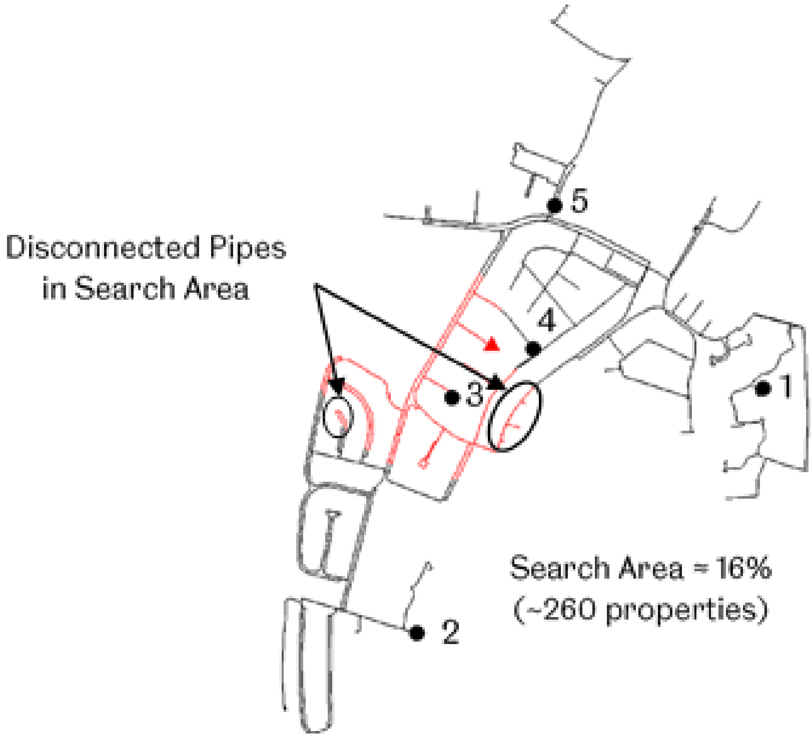


Figure 5-3 *Search area determined for an engineered event using the Euclidean distance function with two disconnected pipes highlighted*

5.3.2. Case Study 5.2: Normalisation of the Pressure Residuals

The process of normalising the pressure residuals prior to performing the SC-IDW interpolation technique was described in section 4.4.1.3. Case study 5.2 provides an example of leak/burst events demonstrating the effect of the normalisation on the performance of the leak/burst localisation technique. The aim of the normalisation step is to account for the change in pressure which occurs due to points which are further away from the inlet of the DMA and will tend to have larger average changes in pressure because of this. In section 4.4.1.3. a simple example was given to demonstrate the effect on the pressure residuals for 3 different leak/burst events modelled in a simple network. Case study 5.2 will demonstrate how the search area changed when the pressure residuals were normalised using several different modelled leak/burst events using a hydraulic model of a real DMA.

To demonstrate the effect of normalising the pressure residuals two versions of the leak/burst localisation technique were developed and used to localise an engineered event. The difference between each of the versions was that the changes in pressure for the engineered event were calculated in a different way. The first version, referred to as the baseline leak/burst localisation technique, used no form of normalisation. The calculated changes in the pressure were determined and used directly in the baseline leak/burst localisation technique. The baseline version of the leak/burst localisation technique is identical to the normalised technique aside from the omission of the normalisation step.

In Figure 5-4 the search area produced by both versions of the leak/burst localisation technique is shown for one engineered event. In this case the larger search area, which also correctly localised the engineered event, was a product of the normalisation procedure which reduced the relative influence of sensors furthest away from the engineered event. Also the relative influence of the two sensors closest to the engineered event, which were sensors 3 and 4 in this case, was adjusted so that the leak/burst event was included in the search area whilst excluding some pipes around sensor 4, which it was desirable to exclude in order to reduce the size of the search area. The pressure differences using both the baseline and the normalised leak/burst localisation technique are given in Table 5-1.

Table 5-1 *Pressure differences for each of the methods of normalising the changes in pressure data for an engineered event*

Leak/Burst Localisation Technique	Sensor 1	Sensor 2	Sensor 3	Sensor 4	Sensor 5
Baseline	0.31	0.23	0.39	0.32	0.33
Normalised	0.11	0.11	0.16	0.14	0.14

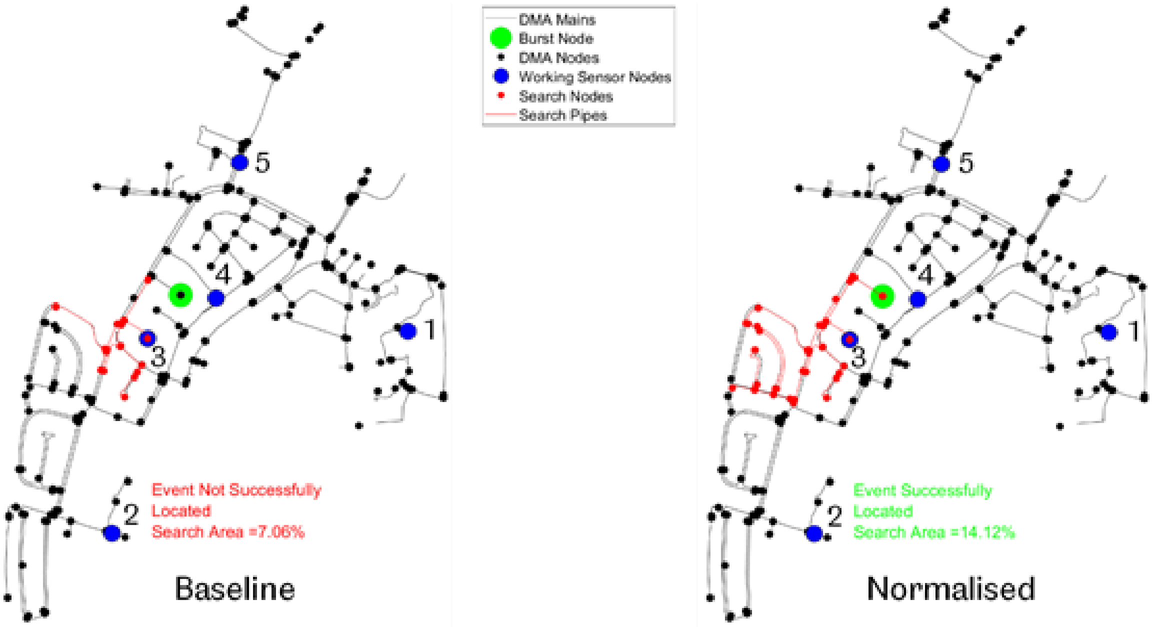


Figure 5-4 *Search areas for an engineered event for both versions of the leak/burst localisation technique*

5.3.3. Case Study 5.3: Engineered Event Localisation Case Study – DMA 123-09

To demonstrate the performance of the leak/burst localisation technique a case study was completed using a set of engineered events conducted in DMA 123-09. The engineered events were localised in an offline fashion using the leak/burst localisation technique and the resulting localisation accuracy for each engineered event was recorded. Engineered event data was used as it contains the natural fluctuations in both pressure and flow which are a normal part of WDS operation to test the leak/burst localisation technique in a more realistic manner than using hydraulic model data. Using engineered event data is a more suitable test for any leak/burst detection or localisation technique as it more closely matches the conditions under which those techniques would need to operate. A benefit of using engineered event data as opposed to data collected during real leak/burst events is that stricter control can be maintained over their characteristics than for naturally occurring leak/burst events. The characteristics of the leak/burst events of interest include the location, the size and the timing. In addition to this there is no need to wait for leak/burst events to spontaneously occur and to search for their location so that the accuracy of the localisation can be confirmed.

5.3.3.1. Engineered Events and Sensor Locations

In DMA 123-09, 11 engineered events were conducted over a three-month span from December 2015 to March 2016. The engineered events were conducted during the night which coincides with the chosen analysis period used by the leak/burst localisation technique. Performing the engineered events during the night also minimised the risk of impacting customer supplies and causing disruption in residential areas. Additionally, the flow rates of the engineered events were selected to prevent issues with customer supplies. The average flow was calculated over a typical week, consisting of 672 flow measurements (one collected every 15 minutes), which was representative of the flow to the DMA prior to the engineered events. The average inflow to the DMA was determined as 8 l/s and the range of engineered event sizes was between 0.5-1.0 l/s giving an approximate normalised leak/burst flow rate of between approximately 6-12% of the average daily DMA inflow. The engineered event flow rate is an important consideration for several reasons. Leak/burst events which are too small or too large will be impossible to localise or can cause severe negative impacts for customers in the DMA. On top of this, larger events are easier to localise but it is desirable to use engineered event sizes which are being targeted by the framework of methods. The locations of the engineered events were selected to give good coverage of the entire DMA to see what effect proximity to the inlet and the connectivity in the area of the engineered event had on the leak/burst localisation performance. A total of 7 sensors were deployed in two phases. During the first phase sensors 1-5 were deployed and then between engineered events 6 and 7 sensors 6 and 7 were added. The sensors were deployed using engineering judgement to ensure even coverage of the DMA and to ensure that any sections of the DMA fed by a single pipe contained at least one pressure sensor. The pressure sensors used were set to record a reading every minute throughout the day.

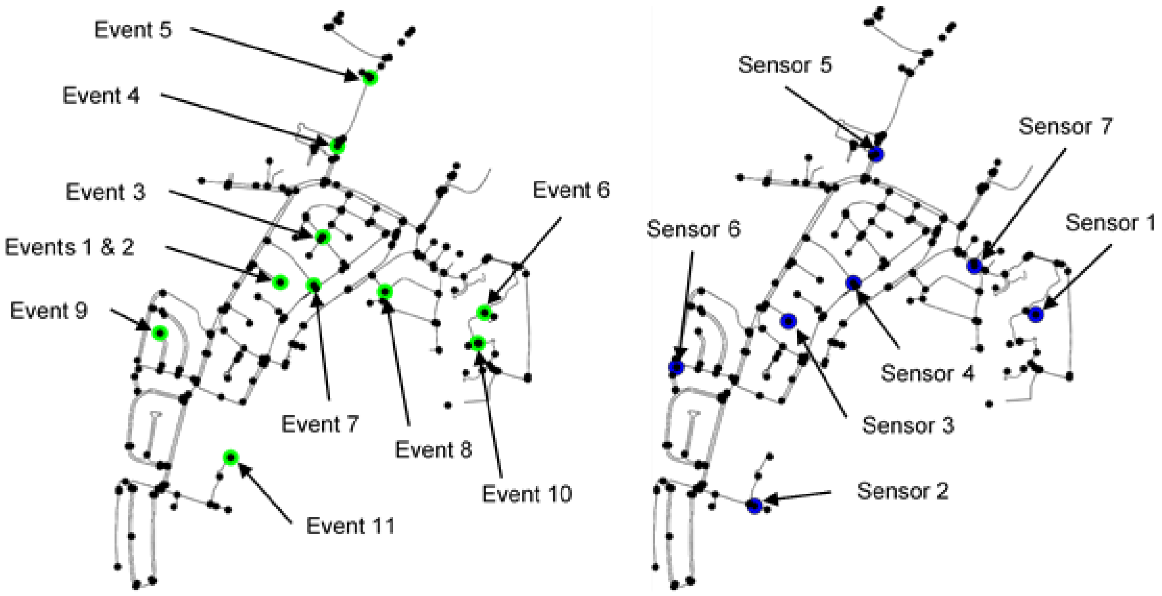


Figure 5-5 *Location of the 11 engineered events (left) and 7 pressure sensors (right)*

5.3.3.2. Leak/Burst Localisation Parameters

To maximise the performance of the leak/burst localisation technique the best values of the leak/burst parameters were determined for the deployed sensor configuration to mimic how they would be determined by the sensor placement technique, even though the deployed sensor configuration is not the optimal one for this purpose. To do this all possible combinations of leak/burst parameters were evaluated. The interpolation exponent was allowed to take integer values between 1 and 30 and the search area threshold could take values between 0.5 and 0.95 in 0.05 increments as was described in section 4.4.2.4.10.

5.3.3.3. Identification of Missing/Faulty Sensor Data

Case study 5.3 was performed retrospectively meaning that all of the engineered events were conducted and the leak/burst localisation technique was performed for all engineered events after the data was collected. Prior to running the leak/burst localisation technique it was necessary to identify pressure sensor data which is either missing or spurious. Taking spurious data first, pressure data can be considered spurious if it does not conform to the normal diurnal pattern which is often seen in pressure data collected from WDSs. In the normal case, assuming that all operating conditions in the WDS are the same, the pressure at a specific point in a DMA will follow a pattern. The pattern will depend upon several factors including whether the DMA pressures are controlled using pressure management valves/pumps or whether there are industrial consumers in it. It is desirable, although not feasible, to consider only one set of operating conditions of the WDS when performing the leak/burst localisation technique. This would ensure that the day on which the leak/burst event occurred will only ever be compared to other days on which the same operating conditions, aside from the presence of the leak/burst event, existed. In actuality, the exact operating conditions can never be known as they vary from day to day, as in the case for customer demand, and often are not measured due to the unfeasibly high cost of doing so. To overcome this issue, the measured pressures were used to infer whether each of the potential previous days was similar to the leak/bust event day or whether there were significant differences, due to any cause, which would preclude its use by the leak/burst localisation technique. Such causes include other leak/burst events, routine maintenance activities, unusually high demand such as on bank holidays, or a rezone within a DMA or upstream of it.

It is important to identify that there is a difference between sensor errors, which cause apparent changes in the pressure, whereas changes in the operating conditions cause actual changes in the pressure in the WDS. Distinguishing between the two is not necessary for the purpose of identifying the days which should not be considered by the leak/burst localisation technique. The way that these changes are manifested within the pressure signals is important as this dictates the means by which they can be identified and hence remove from consideration. Within the pressure data collected from the DMA used in the case studies presented in this section there were several types of common fault with the pressure sensor data which led to the removal of days from consideration by the leak/burst localisation technique. These fault types are described next.

- Pressure spikes

Sometimes erratic changes in the measured pressure can occur which persist for short periods of time before returning to the normal level of pressure. If the spike in pressure was large enough or persisted for multiple time steps then it could also influence the average pressure calculated for a particular day.

- Flatlining

Flatlining refers to the situation where the sensor gets stuck on the same reading for multiple measurements in a row. This type of fault can often be overlooked because each individual pressure measurement appears realistic. As the pressure sensor is providing a faulty reading then some legitimate change in the pressure or another type of fault with the pressure sensor can be missed.

- Instantaneous shifts

In some cases, large and sudden changes in the pressure are seen although the underlying pattern still remains in the data. This can either be due to a legitimate operation, for example a re-zone or operation of a pumping station or can be due to a fault with the sensor. Each of the common types of sensor faults are demonstrated in Figure 5-6 using plots of single pressure sensor data collected from DMA 123-09. This list of common sensor faults is not exhaustive and only the types found in the pressure data from DMA 123-09 are shown.

Missing pressure sensor data is easier to identify and in the offline context of this case study a simpler matter to deal with. To deal with missing sensor data there were two scenarios of interest. Firstly, in the case where some sensor data was missing on the day when the leak/burst event occurred. In this case the remaining pressure data was used to determine the average pressure for each sensor. As the night time was being used as the analysis window by the leak/burst localisation technique the pressure in the analysis window did not vary significantly as a function of the reading which is selected which means that considering only a fraction of the analysis window will not have a significant impact upon the average pressure due to the missing data. For days which fall in the previous days data set if any data is missing from the analysis window then the day is marked as incomplete and will be ignored by the leak/burst localisation technique. For the purposes of this case study a manual inspection of the pressure data from all sensors was conducted to identify any days containing at least one of these faults. For each sensor any day containing at least one fault was tagged so that it could be excluded during the formation of the previous days data set prior to performing the leak/burst localisation technique.

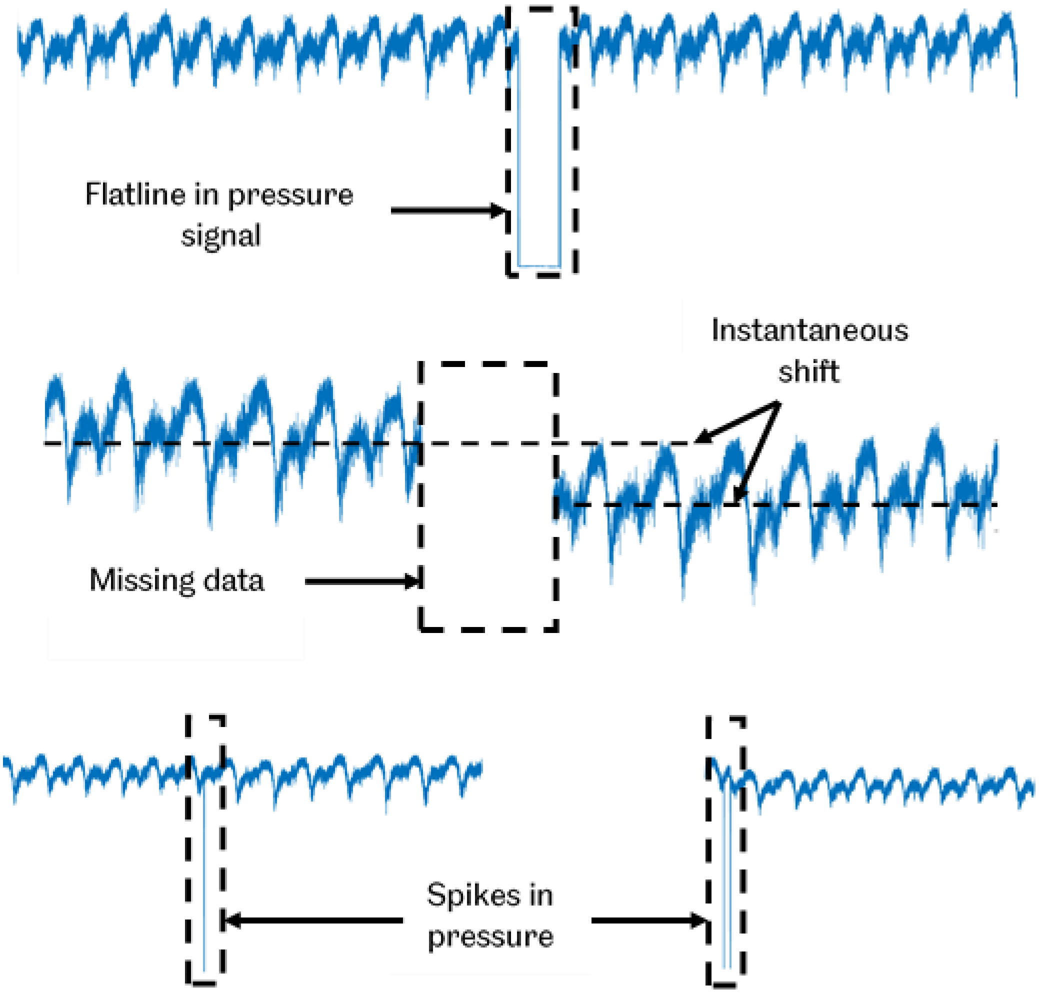


Figure 5-6 *Common sensor fault types in the pressure data collected from DMA 123-09 collected on days with no engineered events being conducted*

5.3.3.4. Summary of Results

The leak/burst localisation results for two versions of the leak/burst localisation techniques are given in Table 5-2. For each engineered event, the number of deployed and working sensors are given along with the size of the search area. Each engineered event which was correctly localised is denoted by the use of bold text. At the bottom of Table 5-2 the average size of the search area, considering all engineered events, is given.

Table 5-2 *Summary of the localisation performance for 11 engineered events conducted in DMA 123-09 using two different versions of the leak/burst localisation technique. For each engineered event, the size of the search area, as a proportion of the size of the DMA, is given. Results in bold were correctly localised*

Event Number	Event Date	Working Sensor Count (Total installed)	Baseline	Normalised
1	18/12/2015	5 (5)	7	**14**
2	23/12/2015	4 (5)	8	**15**
3	07/01/2016	5 (5)	**61**	**52**
4	12/01/2016	5 (5)	**51**	**55**
5	20/01/2016	5 (5)	**14**	**26**
6	22/01/2016	5 (5)	**10**	**11**
7	16/02/2016	7 (7)	9	10
8	19/02/2016	7 (7)	9	10
9	24/02/2016	7 (7)	**8**	**10**
10	01/03/2016	6 (7)	2	4
11	08/03/2016	6 (7)	**14**	**15**
Average (% DMA)	-	-	17.5	20.2

The main highlight of the results in Table 5-2 is that the normalised leak/burst localisation technique was able to correctly localise more engineered events than the baseline leak/burst localisation technique. In addition, there is no significant difference between the average size of the search area when only the correctly localised events are considered which indicates that the improved localisation performance attained by including the normalisation step in the leak/burst localisation technique is not achieved because larger search areas are produced. Indeed, even for the incorrectly localised engineered events there is only a negligible increase of around 1% (7% for the baseline and 8% for the normalised) in the average size of the search area when considered in relation to the size of the DMA. This means that the improvement in performance was caused by the search area being more focussed around the engineered event locations thereby more accurately localising them.

5.4. Sensor Placement Technique

5.4.1. Case Study 5.4: Sensitivity Analysis for Investigating the Leak/Burst Localisation Parameters Effect on Performance

As was outlined in section 4.4.2.4.6, a procedure for automatically determining the best values for the interpolation exponent and the search area threshold was developed as part of the sensor placement technique. This procedure is critical because of the strong dependence of the performance of the leak/burst localisation technique on the values of both leak/burst localisation parameters. This case study demonstrates the need for and benefit of using the automatic procedure rather than determining the values of the leak/burst localisation parameters manually. This case study served several important purposes which are listed below:

- To demonstrate the interaction of the two leak/burst localisation parameters to identify how each affected the leak/burst localisation performance, measured by the average size of the search area and the number of leak/burst events with multiple search areas, of the sensor placement technique. This was to determine whether the best combination of leak/burst localisation parameters was specific to each network.
- Additionally, the effect of the number of sensors on the best values of leak/burst localisation parameters was investigated to determine whether the same leak/burst localisation parameter values could be used independent of the number of sensors being considered.
- To establish a baseline Pareto front against which the automatic parameter selection technique could be compared to ensure it performed well.

This case study was a sensitivity analysis where the values of both leak/burst localisation parameters were varied independently to examine what effect they had on the performance of the sensor placement technique. To do this, the two simple verification hydraulic models were used because they were smaller than DMA hydraulic models. This allowed multiple runs (one for each combination of leak/burst localisation parameters) of the sensor placement technique to be completed in a reasonable time. To limit the number of runs of the sensor placement technique, 30 combinations of leak/burst localisation parameters were used. The 30 combinations of parameters were obtained by using five different values (0.5 to 0.9 in 0.1 increments) for the search area threshold and six different values (integers ranging from 1-6) for the interpolation exponent. The interpolation exponent values were taken from the literature of previous applications of IDW interpolation (Zimmerman, 1999). The search area threshold values were selected to cover the practical range of values which are between 0.5 and 1 which was explained in section 4.4.2.4.10. For each combination of parameters, a single run of the sensor placement technique was completed and the resulting leak/burst localisation performance determined, using the average size of the search area.

Figures 5-7 and 5-8 show the average leak/burst localisation performance achieved for each combination of parameters for both simple verification hydraulic models. The sensitivity analysis was performed for both simple verification hydraulic models to see if there were any differences between the leak/burst localisation performance for different combinations of leak/burst localisation parameters for both verification hydraulic models. In the cases presented in Figure 5-7 and Figure 5-8, 10 sensors were considered. In Figure 5-7 the leak/burst localisation performance for the 30 different parameters combinations is depicted by each bar for the dendritic verification hydraulic model. The height of each bar (as well as the colour) is proportional to the leak/bust localisation performance and the shorter the bar the smaller the average search area and the better the leak/burst localisation performance. The interpolation exponent varies along the left horizontal axis and the search area threshold varies along the right horizontal axis.

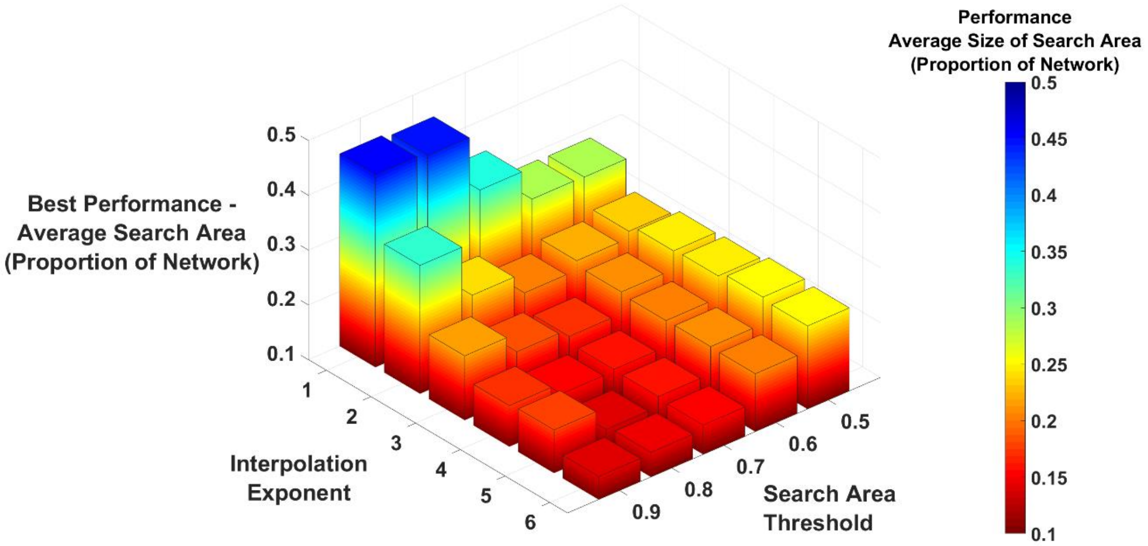


Figure 5-7 *Average size of the search area for the best sensor configurations for the dendritic network with varying interpolation exponent and search area threshold*

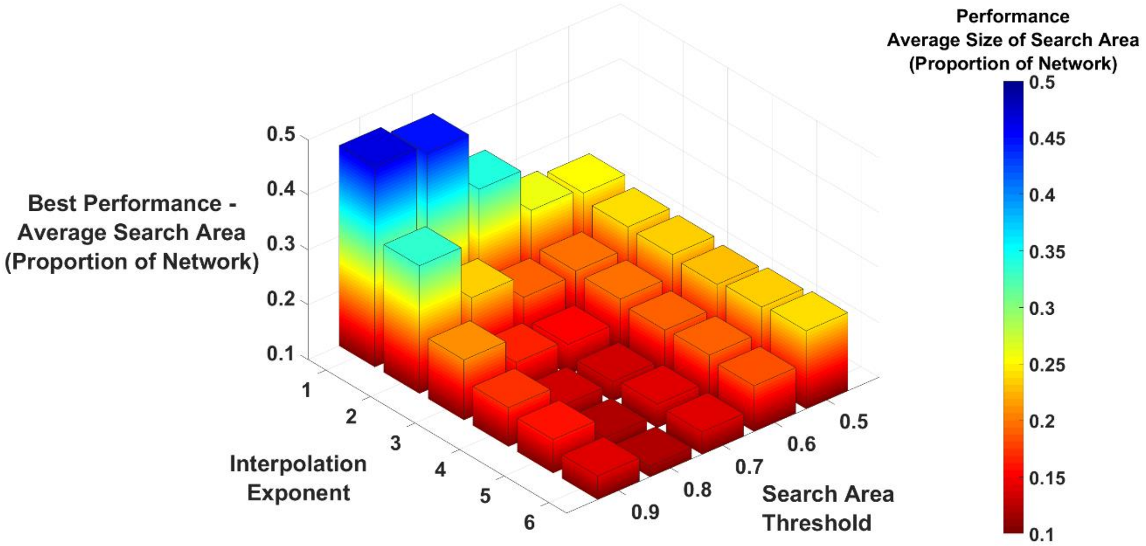


Figure 5-8 *Average size of the search area for the best sensor configurations for the looped network with varying interpolation exponent and search area threshold*

For the purposes of this case study, which was to demonstrate the variation in the performance of the leak/burst localisation technique for different leak/burst localisation parameter combinations, the values of the leak/burst localisation parameters were very influential. From the results in Figures 5-7 several important trends were observed. Firstly, and most importantly, is the variation between the leak/burst localisation performance achieved by using the best combinations of parameters when compared to the worst combinations of parameters. The best combination of parameters was 5 for the interpolation exponent and a search area threshold of 0.8 which gave an average search area of 13.9% of the DMA. In contrast to this, the worst performing combination of parameters, which was an interpolation exponent of 1 and a search area threshold of 0.9, produced an average search area of 45.1%. This demonstrates the importance of choosing a suitable combination of parameters for the leak/burst localisation technique to perform well. Another important point is that selecting leak/burst localisation parameter combinations which were near to (but not the optimal) combination did not result in a severe loss of leak/burst localisation performance. However, in order for this to be confirmed for each different network a full sensitivity analysis would be required by which point the best combination of leak/burst localisation parameters would be known anyway. This is particularly problematic for larger DMAs, where many leak/burst event locations usually must be considered. For this reason, using a heuristic approach of selecting parameter combinations which appear to work well for multiple networks was not considered suitable. The optimal combination of leak/burst localisation parameters should be determined for each network along with the sensor configurations to ensure that, regardless of the network being considered, the best combination of parameters is always identified.

Several interesting phenomena were also seen by varying one parameter and keeping the other fixed. For example, when 6 was selected as the value of interpolation exponent, increasing the search area threshold tended to produce better leak/burst localisation performance. In contrast to this, when 1 was selected as the value of interpolation exponent increasing the search area threshold tended to produce worse leak/burst localisation performance which was the opposite of the trend when the value of 6 was used. For fixed values of search area threshold some interesting relationships were seen for different values of interpolation exponent. By selecting a fixed value of search area threshold equal to 0.9 a steep gradient in the leak/burst localisation performance was found. Increasing the interpolation exponent tended to significantly improve the leak/burst localisation performance. The relationship was approximately exponential and the effect of increasing the interpolation exponent tended to reduce as higher values were used. Considering a search area threshold of 0.5 the leak/burst localisation performance was not significantly affected by varying the leak/burst interpolation exponent. The results for the looped verification hydraulic model are shown in Figure 5-8. There were no significant differences between the dendritic and the looped verification hydraulic models. The only difference is that a slightly different combination of parameters gave the best leak/burst localisation performance. For the looped verification hydraulic model the best interpolation exponent was 6 and the best search area threshold was 0.8 giving an average search area of 12.0% of the DMA. The worst performing combination of parameters produced an average search area of 46.5% of the DMA.

The same procedure was also conducted considering the number of search areas which is the secondary performance metric being used by the sensor placement technique. In Figure 5-9 and Figure 5-10 the number of leak/burst events with multiple search areas was determined for each leak/burst localisation parameter combination and was plotted in the same way as for the average size of the search area considering the case with 10 sensors being used.

It is clear that the parameter combination had a strong effect on the proportion of leak/burst events with multiple search areas. When the combination of parameters was well selected the problem of producing large numbers of leak/burst events with multiple search areas was largely overcome however a poorly chosen combination of parameters led to more than 50% of leak/burst events producing multiple search areas. By comparing the performance of the different parameter combinations using both performance metrics, the variation of the leak/burst localisation performance behaved very differently. Small changes in the parameters tended to produce small changes in the size of the average search area whereas the proportion of leak/burst events with multiple search areas varied drastically even for small changes in only one parameter. Therefore, a key problem is identifying a combination of parameters which simultaneously minimises the number of leak/burst events with multiple search areas and produces small search areas for as many leak/burst events as possible. The same behaviour was observed for the looped case study network in Figure 5-7.

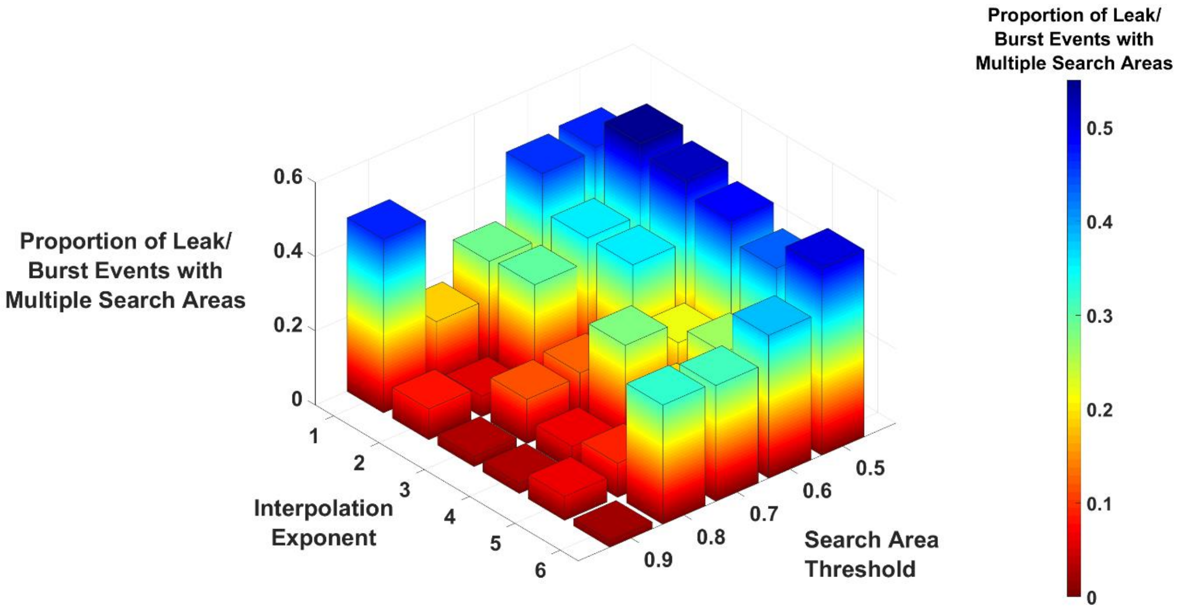


Figure 5-9 *Proportion of leak/burst events with multiple search areas for the best sensor configurations for the dendritic verification hydraulic model with varying interpolation exponent and search area threshold*

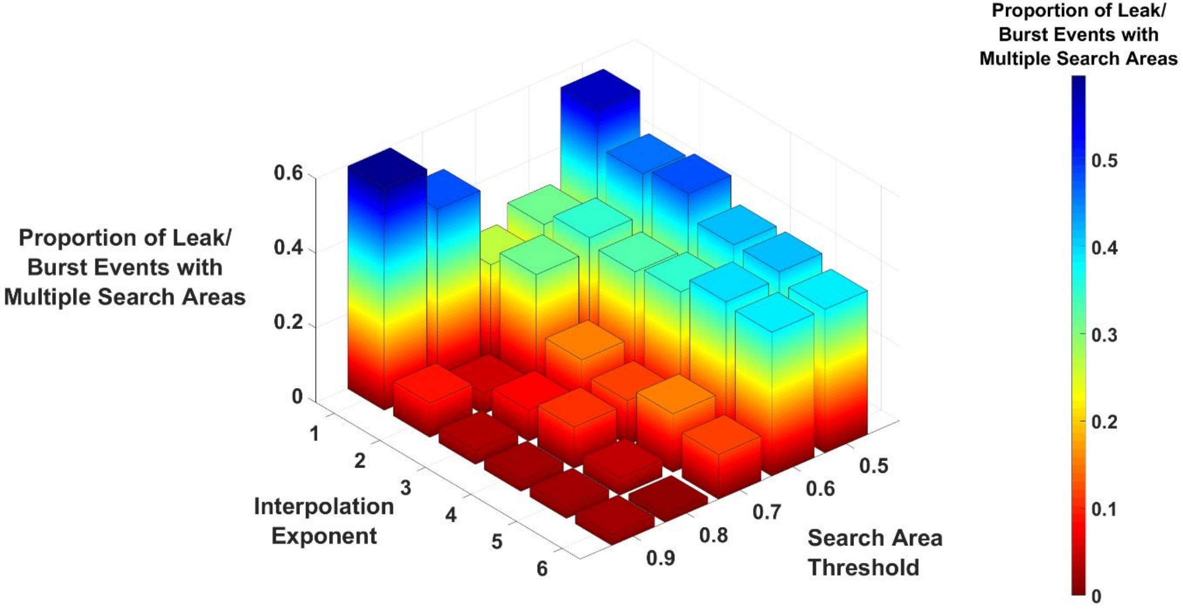


Figure 5-10 *Proportion of leak/burst events with multiple search areas for the best sensor configurations for the looped verification hydraulic model with varying interpolation exponent and search area threshold*

So far, only cases with 10 sensors have been considered which means that a key variable has been ignored. The case study was therefore expanded to include all numbers of sensors which are considered by the sensor placement technique, which is between 3 and 10 sensors as was described in section 4.4.2.4.10. In Figures 5-11 and 5-12 the average size of the search area is shown for both of the simple verification hydraulic models for all combinations of interpolation exponent, search area threshold and number of sensors. In each figure the colour of the circle denotes the average size of the search area. The red and orange circles denote the combination of parameters and number of sensors which give the best localisation performance. For each number of sensors, the best combination of parameters is highlighted using a black circle. From Figures 5-11 and 5-12 it was seen that the best parameter combinations were very similar but did vary as the number of sensors did. This indicates that selecting fixed values of the parameters for different numbers of sensors will not produce the best leak/burst localisation performance when using the average size of the search area as the performance metric.

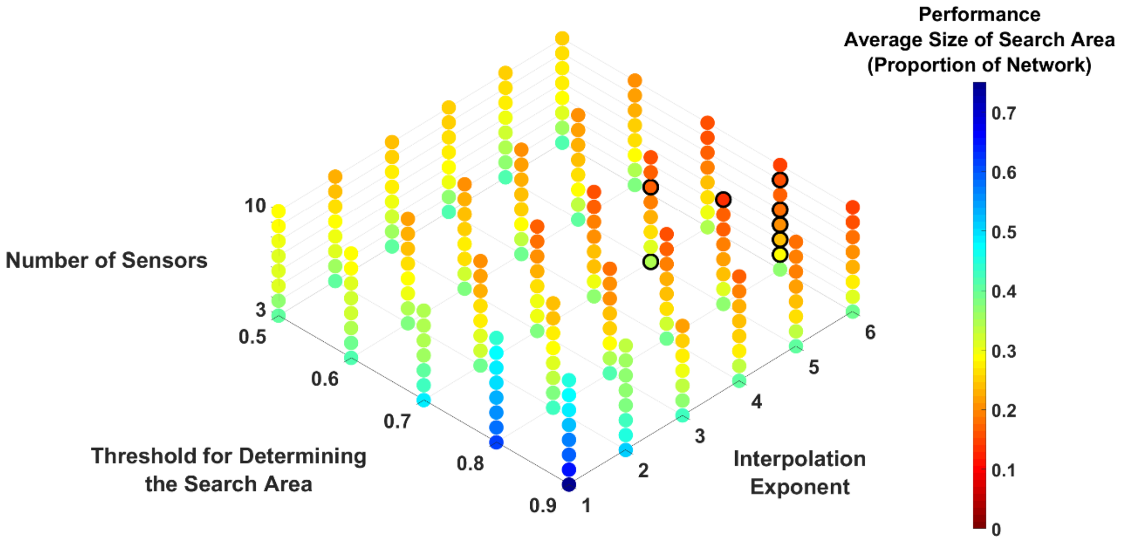


Figure 5-11 *Average size of the search area for all sensor configurations for the dendritic network with varying interpolation exponent and search area threshold. The best performing parameter combinations have been highlighted in black for each number of sensors*

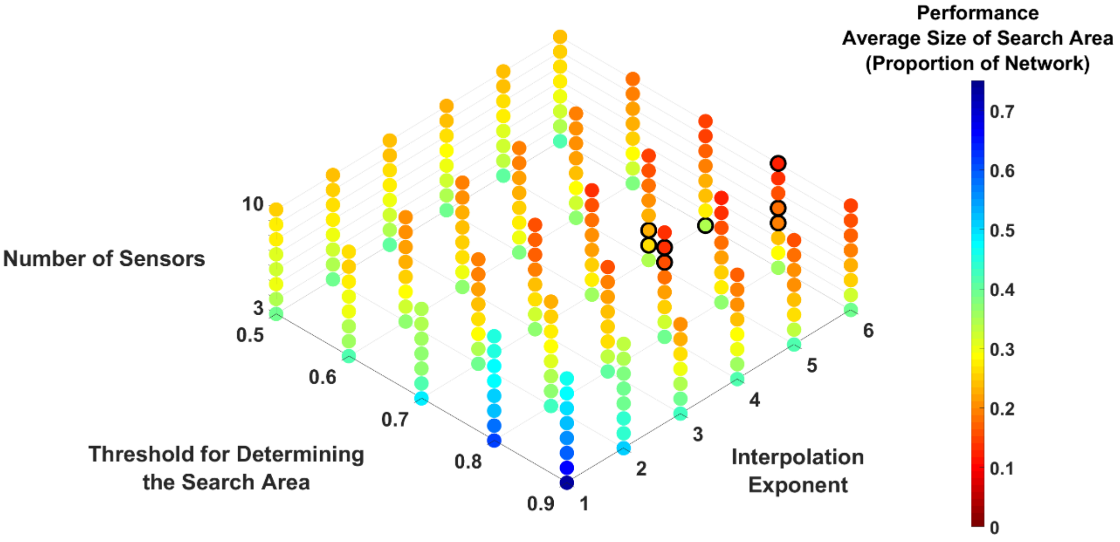


Figure 5-12 *Average size of the search area for all sensor configurations for the looped network with varying interpolation exponent and search area threshold. The best performing parameter combinations have been highlighted in black for each number of sensors*

In Figure 5-13 and Figure 5-14 the number of leak/burst events with multiple search areas was plotted for all combinations of leak/burst localisation parameters and numbers of sensors for both verification hydraulic models. Combining the results in Figures 5-11, 5-12, 5-13 and 5-14 the need for including the penalty for producing multiple search areas in the sensor placement objective function is demonstrated. Even when good leak/burst localisation performance is achieved according to the average size of the search area many leak/burst events can have multiple search areas which is not desirable. By penalising the leak/burst localisation performance in proportion to the number of leak/burst events with multiple search areas the sensor placement technique can be steered towards sensor configurations for which this problem is minimised.

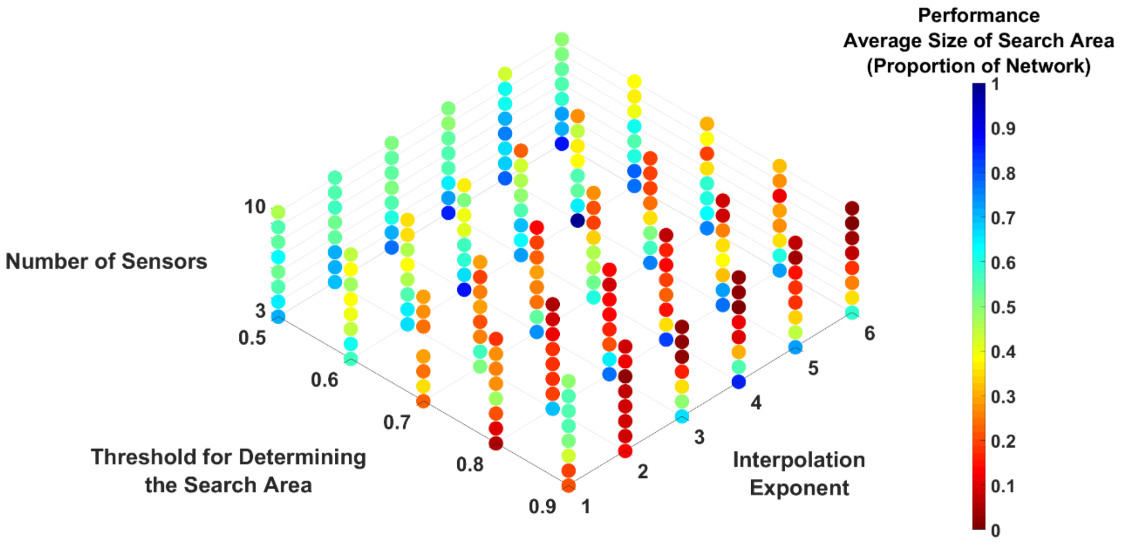


Figure 5-13 *Number of leak/burst events with multiple search areas for all sensor configurations for the dendritic verification hydraulic model with varying interpolation exponent, search area threshold and number of sensors*

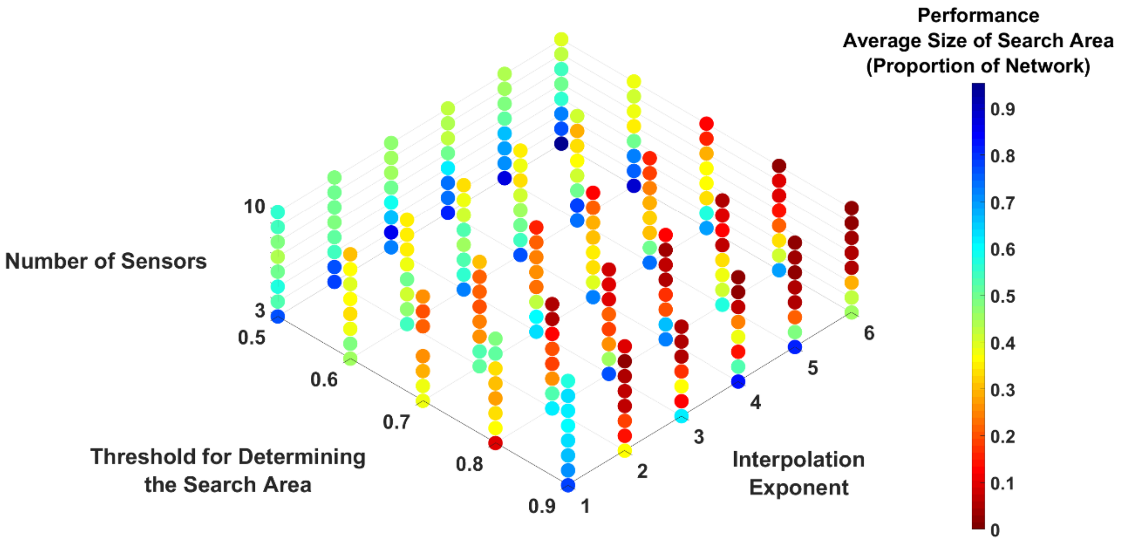


Figure 5-14 *Number of leak/burst events with multiple search areas for all sensor configurations for the looped verification hydraulic model with varying interpolation exponent, search area threshold and number of sensors*

5.4.2. Case Study 5.5: Testing the Automatic Parameter Selection Technique

Case study 5.5 demonstrates the ability of the automatic parameter selection technique to identify the best combination of leak/burst parameters when used as part of the sensor placement technique. From the 30 parameter combinations which were used to generate the results for case study 5.4 the best leak/burst localisation parameter combinations were extracted for each number of sensors to give a baseline Pareto front. The baseline Pareto front is used to determine whether the best-known parameter combinations have been achieved by the automatic parameter selection technique. In the ideal case exactly the same parameter values would be identified by the automatic parameter selection technique as were identified when the 30 individual sensor placement runs were performed. Given that the GALAXY MOEA, which is being used to determine the optimal sensor configurations, is a heuristic its ability to determine the same optimal solutions is not guaranteed. By combining the search for the optimal sensor configurations with the search for the optimal combination of leak/burst localisation parameters the shape and size of the search space is modified. This means that using the GALAXY MOEA (or any GA) for both will not guarantee convergence to the same solutions. The purpose of this case study is to test the ability of the GALAXY MOEA to achieve this.

In Figures 5-15 and Figure 5-16 the comparison between the baseline Pareto front and the Pareto fronts obtained by using the automatic parameter selection technique are shown for both simple verification hydraulic models. The best leak/burst localisation performance achieved in case study 5.4 are denoted by the red crosses and the best values of the parameters are given in red text. The search area threshold is given first followed by the interpolation exponent. The best leak/burst localisation performance and the corresponding parameter values are shown in black for the automatic parameter selection technique. For both simple verification hydraulic models, the same or a very similar level of leak/burst localisation performance was achieved by both methods of determining the optimal combination of parameters.

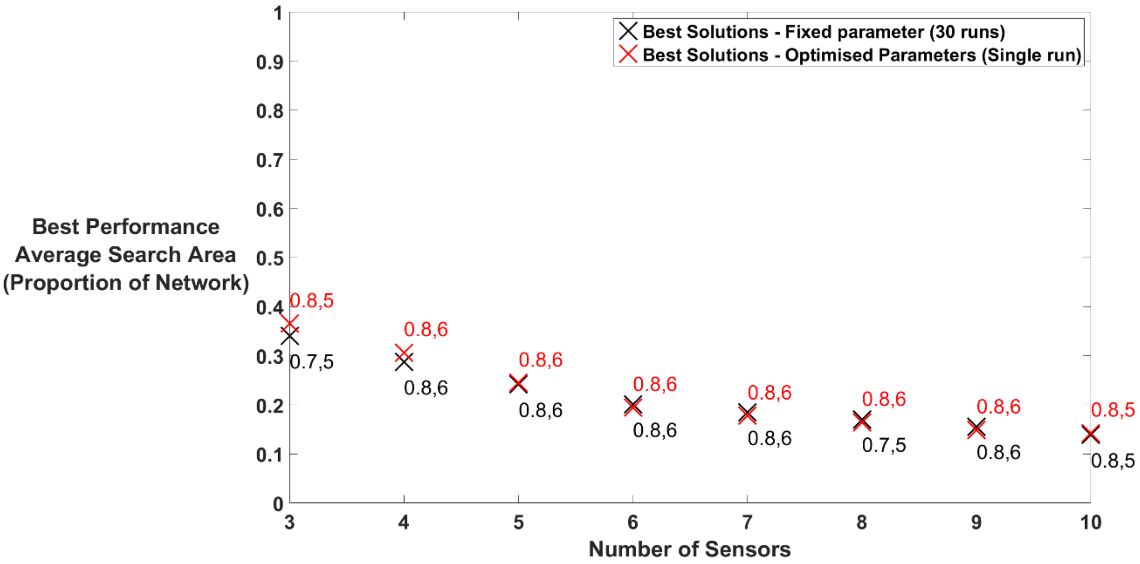


__Figure 5-15__ Best set of parameters determined using 30 individual sensor placement runs with fixed parameters (red crosses) and sensor configurations included in the optimisation (black crosses) for the dendritic verification hydraulic model

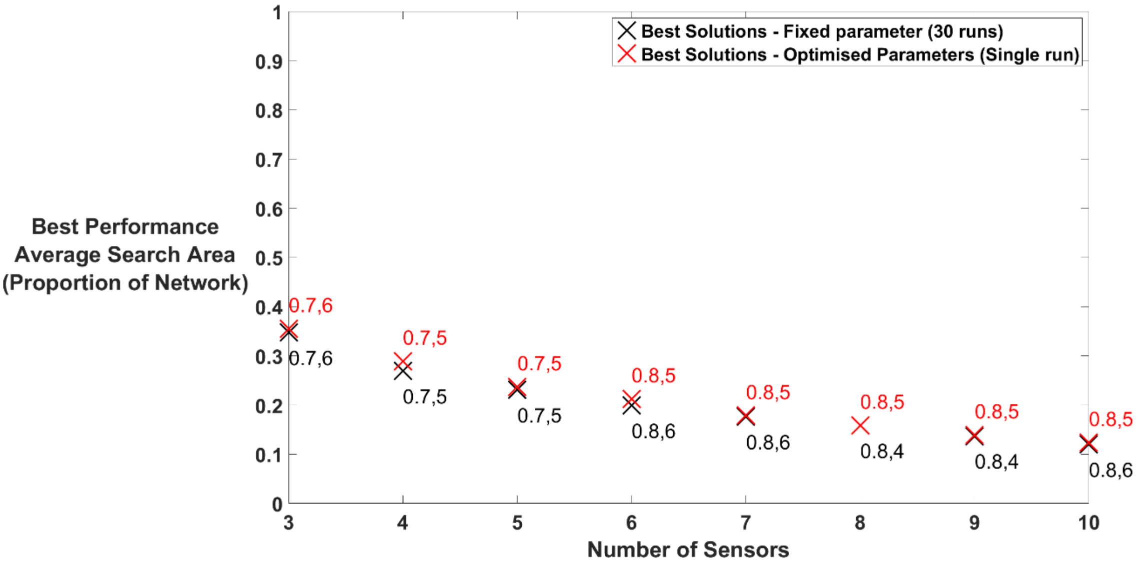


__Figure 5-16__ Best set of parameters determined using 30 individual sensor placement runs with fixed parameters (red crosses) and sensor configurations included in the optimisation (black crosses) for the looped verification hydraulic model

For the dendritic verification hydraulic model, for which the results are shown in Figure 5-15, for the cases with greater than 4 sensors the leak/burst localisation performance was almost identical. For the cases with more than 4 sensors the differences between the leak/burst localisation performance when using the automatic parameter selection technique and the manual sensitivity analysis were less than 3% of the overall size of the dendritic case study network. Slightly different parameter combinations were also identified by the automatic parameter selection technique for the cases with 3 and 8 sensors although for the remaining numbers of sensors identical parameter combinations were determined. Given the heuristic nature of the GALAXY MOEA, which was used to determine both the baseline Pareto front and the Pareto front with the parameters simultaneously being determined, this is not considered to be a serious problem. Given that the same number of function evaluations were used in both cases and that simultaneously determining the parameters changes the size and shape of the search space this was not unexpected. In addition to this because the sensor configurations could differ between the baseline and the automatically selected parameters it was possible for the sensor configurations to contribute to the performance differences which were seen. In Table 5-3 the sensor configurations for both the baseline (left) and the automatically selected parameters (right) are shown for the dendritic case study network.

Table 5-3 *Optimal sensor configurations for the dendritic verification hydraulic model for 30 manual sensor placement runs (left) and the single run with optimised parameters (right)*

	Dendritic - Manual								**Dendritic - Optimised**							
Number of Sensors	3	4	5	6	7	8	9	10	3	4	5	6	7	8	9	10
Sensor Indices	13	48	11	11	11	17	4	4	52	49	52	17	17	17	17	17
	38	60	51	17	25	26	10	10	65	52	65	30	24	26	26	26
	105	86	81	25	79	52	38	38	79	65	86	53	26	30	52	39
		105	90	66	81	66	58	57		79	90	65	49	39	65	53
			104	90	87	72	61	61			104	90	53	49	79	65
				103	90	90	68	68				104	65	53	86	79
					103	103	99	79					104	65	90	90
						106	103	99						79	103	97
							106	102							105	99
								106								105

Whilst many of the individual sensor locations were shared between the two there were some differences which indicates that, along with the slightly different parameters, the sensor configurations contributed to the differences in leak/burst localisation performance which were found. Given that the agreement between the baseline Pareto front and the Pareto front with automatically determined parameters was very good, and that the parameter combinations were also very similar, the performance of the automatic parameter selection procedure was sufficient to be used as part of the sensor placement technique. The finding was very similar for the looped verification hydraulic model, for which the results are shown in Figure 5-16. The maximum difference between the leak/burst localisation performance for the baseline Pareto front and the Pareto front for the automatically determined parameters was 2% of the size of the looped case study network. The same pattern of selecting different sensor configurations for all numbers of sensors was seen for the looped case study network and is shown in Table 5-4.

Table 5-4 *Optimal sensor configurations for the looped verification hydraulic model for 30 manual sensor placement runs (left) and the single run with optimised parameters (right)*

	Looped - Manual								**Looped - Optimised**							
Number of Sensors	3	4	5	6	7	8	9	10	3	4	5	6	7	8	9	10
Sensor Indices	3	3	13	26	26	26	16	16	13	13	13	26	26	26	15	26
	91	38	38	36	47	36	26	28	60	51	48	36	36	36	26	34
	98	72	48	38	51	51	52	47	105	59	51	47	46	52	36	36
		106	72	52	58	52	58	51		105	59	51	51	59	52	51
			105	66	59	58	60	52			105	59	58	60	58	52
				104	69	59	66	58				104	59	66	60	58
					104	66	68	59					104	88	66	59
						104	71	66						104	71	66
							104	69							104	99
								104								104

The performance of the automatic parameter selection technique means that a higher number of parameter combinations can be considered by the sensor placement technique as only one run of the sensor placement technique is needed. Further improvements in performance may be achieved for the two simple verification hydraulic models by increasing the range of allowable interpolation exponents and number of values that the search area threshold can take. Therefore, for future applications of the sensor placement, particularly for real DMAs, it is recommended that the maximum allowable value for the interpolation exponent is substantially increased and that the increment of the search area threshold is reduced so that more values are considered.

5.4.3. Case Study 5.6: Testing the Penalty for Multiple Search Areas

As was highlighted by the case study in case study 5.4, which investigated the leak/burst localisation parameters, careful selection of the parameters could not completely overcome the problem of producing multiple search areas for all leak/burst events. Even the best combination of leak/burst localisation parameters still produced several leak/burst events with multiple search areas. Further complicating the problem is that the leak/burst localisation parameter combinations which produced the smallest number of leak/burst events with multiple search areas did not necessarily produce the smallest average search area size. To solve this problem a penalty was introduced in the sensor placement objective function to automatically steer the sensor placement technique towards sensor configurations which produced good performance considering both performance metrics.

Presented in this section is a case study to demonstrate the performance improvements, in terms of the number of leak/burst events with multiple search area, which were achieved by introducing the penalty to the sensor placement objective function. The case study compares the performance of the sensor placement technique without the penalty applied with the performance achieved when the penalty was included. To do this both simple verification hydraulic models were used. Sensor placement runs using the penalised objective function and the unpenalised objective function were completed so that the penalty was the only difference between the two scenarios. For each of the scenarios the average size of the search area and the number of leak/burst events with multiple search areas were calculated once the optimal sensor configurations were determined. Figure 5-17 and Figure 5-18 show the performance for the penalised and unpenalised objective functions for the dendritic and looped verification hydraulic models, respectively.

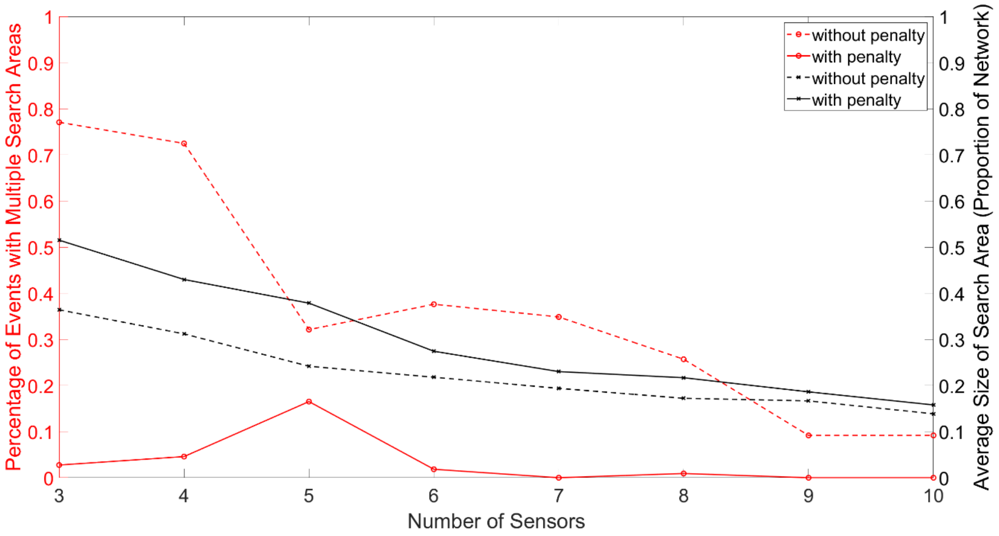


Figure 5-17 *Comparison of the proportion of leak/burst events with multiple search areas (left vertical axis) and the average size of the search areas (right vertical axis) for the penalised and unpenalised objective functions for the dendritic verification hydraulic model*

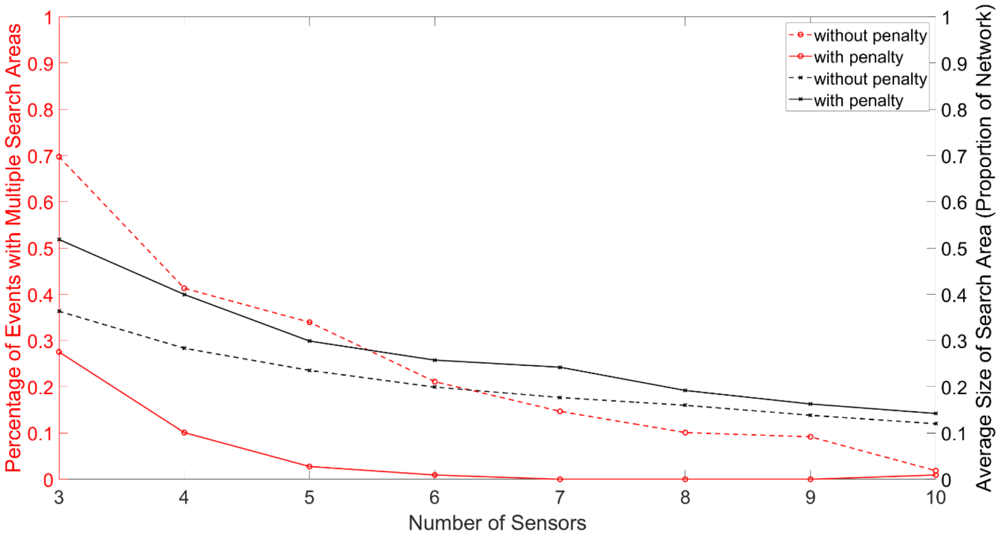


Figure 5-18 *Comparison of the proportion of leak/burst events with multiple search areas (left vertical axis) and the average size of the search areas (right vertical axis) for the penalised and unpenalised objective functions for the looped verification hydraulic model*

In both cases the number of leak/burst events with multiple search areas is shown in red on the left vertical axis. The average size of the search areas considering all modelled leak/burst events are shown in black and correspond to the values on the right vertical axis. For the scenarios presented in this section a fixed size of emitter coefficient was used for all leak/burst event locations and the best-known combination of parameters was used for each.

Considering the results in Figure 5-17, related to the dendritic verification hydraulic model, the most obvious effect of the penalty is that the number of leak/burst events with multiple search areas was drastically reduced for all numbers of sensors considered. For the dendritic case study network the number of leak/burst events with multiple search areas was reduced by between 74% (3 sensors) and 9% (10 sensors). The reduction in the number of leak/burst events with multiple search areas was loosely related to the number of sensors used. For all cases with greater than 6 sensors being considered only one leak/burst event produced multiple search areas. This demonstrates the ability of the penalty to entirely overcome the problem of producing multiple search areas although this is dependent upon the number of sensors. The average size of the search area was increased by including the penalty within the objective function. The size of the increase in the search area was also linked to the number of sensors which are considered with a smaller increase in the average size of the search area when higher numbers of sensors are considered.

A similar picture was seen for the looped case study network, for which the results are shown in Figure 5-18. A reduction of between 42% (3 sensors) and 1% (10 sensors) in the number of leak/burst events with multiple search areas was achieved for all numbers of sensors considered. In the same way as for the dendritic case study network an increase in the average size of the search area was found for all numbers of sensors. The extent of both the reduction in the number of leak/burst events with multiple search areas and the increase in the average size of the search area was related to the number of sensors which were considered.

5.4.4. Case Study 5.7: Measuring the Size of the Search Area

Case study 5.7 demonstrates the effect of measuring the search area for each leak/burst event using the length of pipes in the search area as opposed to the number of junctions in the hydraulic model. As was stated in section 4.4.2.4.8 the reason for using the length of pipes as opposed to the number of junctions was to better reflect the situation of finding the exact location of leak/burst events in real water distribution systems. The time taken to precisely locate (or pinpoint) a leak/burst event will be a function of the total length of pipes which must be examined using hardware-based techniques, such as leak noise correlators. Using the length of pipes rather than the number of junctions, which are not equivalent between hydraulic models and real WDSs, is more representative of the situation in real WDS. Junctions in real WDSs typically refer to the vertex between two or more pipes whereas in hydraulic models junctions can be included to represent customer demands to connect valves, pumps or other pipe fittings to pipes. The purpose of modifying the measurement of the search area was not to reduce the average size of the search area when hydraulic model data is considered but to force the sensor placement technique to select sensor configurations which perform well on real water distribution systems as well as in hydraulic models. Another benefit of this approach is that performance achieved using real leak/burst events can be compared to the performance achieved by the sensor placement technique.

A comparison between the sensor placement performance using the number of junctions and the length of pipes was completed. The sensor placement performance for both search area measures was broadly similar. On average the performance should not differ significantly but the chosen sensor locations should not tend to favour areas of high junction density. This was a known problem of using the junction count to evaluate the leak/burst localisation performance. In Figure 5-19 and Figure 5-20 the sensor placement performance for the dendritic and looped verification hydraulic models, respectively, are shown. For both verification hydraulic models, the sensor placement performance was shown to differ by between 0.5 and 5% of the total size of the networks. This is not considered significant in terms of the overall size of both networks and in all cases, aside from the case with 3 sensors for the dendritic case study network, the performance was slightly improved by using the pipe length as opposed to the junction count.

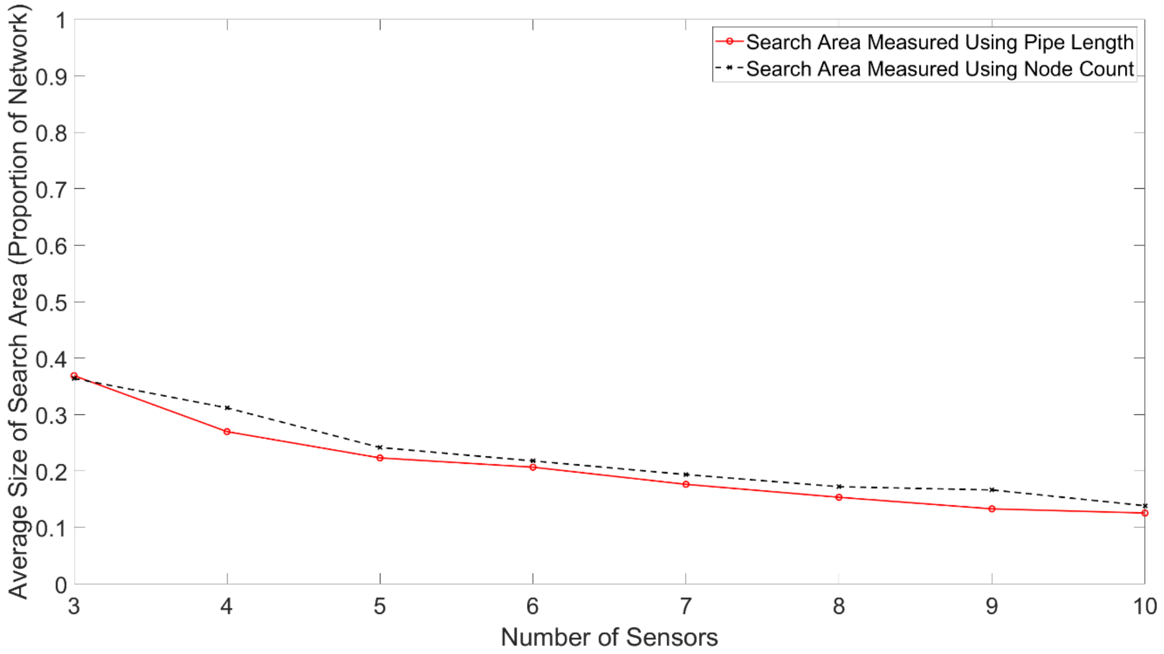


Figure 5-19 *Comparison of the sensor placement performance for both search area measures for the dendritic verification hydraulic model*

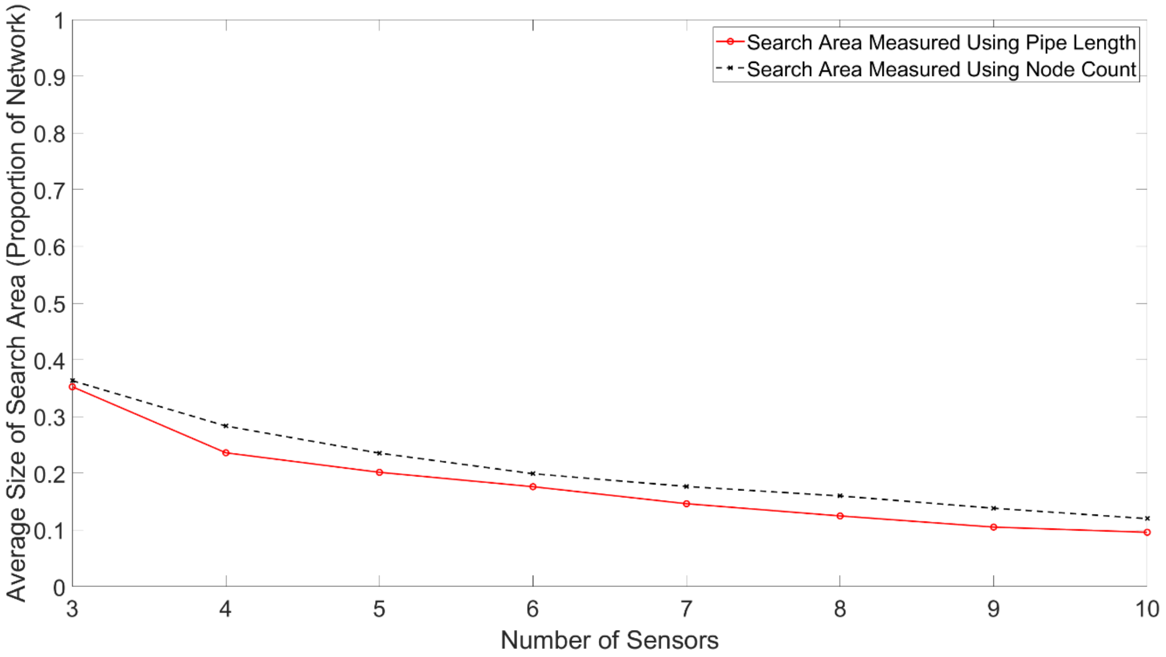


Figure 5-20 *Comparison of the sensor placement performance for both search area measures for the looped verification hydraulic model*

5.4.5. Case Study 5.8: Automatic determination of leak/burst event sizes

The procedure for automatically determining the leak/burst event sizes to be used by the sensor placement technique was described in section 4.4.2.3.6. Case study 5.8 describes the application of the procedure for automatically determining the leak/burst event sizes to the dendritic verification hydraulic model. In particular the selection of two leak/burst event size parameters, namely the event size increment and the maximum allowable flow increase, were examined to determine how the interaction between the two affects the validity of the leak/burst event sizes. The results in this case study demonstrate how, as part of the procedure for automatically determining the leak/burst event sizes at all leak/burst event locations, the applicability of the framework of methods to a particular DMA can be assessed. This provides a means of identifying DMAs which are not sensitive to leak/burst events and should not be considered for sensor placement.

The event size increment governs how quickly the leak/burst event sizes are increased to determine the maximum and minimum leak/burst event size for each leak/burst event location. Using a smaller value more accurately determines the leak/burst event sizes but this comes at the cost of computational time because it increases the number of model runs required. The maximum allowable flow increase sets a threshold for the maximum leak/burst event sizes and must be selected to align with the strategic goals of a water company. To examine the combined effect of the leak/burst event size parameters 10 different values of each leak/burst event size parameter, to give 100 combinations in total. Each combination of parameters was used to perform a run of the procedure for automatically determining the leak/burst event sizes. To determine whether valid leak/burst event sizes were produced for each leak/burst event size parameter combination, the maximum and minimum leak/burst event sizes were determined for all leak/burst event locations. If, for any leak/burst event location, the maximum leak/burst event size was smaller than the minimum leak/burst event size then the combination of parameters was considered invalid. Figure 5-21 shows the validity of the 100 leak/burst event size parameter combinations for the dendritic verification hydraulic model. For each combination of parameters that produced valid results a black dot was plotted. Conversely, when the results were invalid, a red circle was plotted. The boundary between the valid and invalid results is not linear against either of the parameters in Figure 5-21.

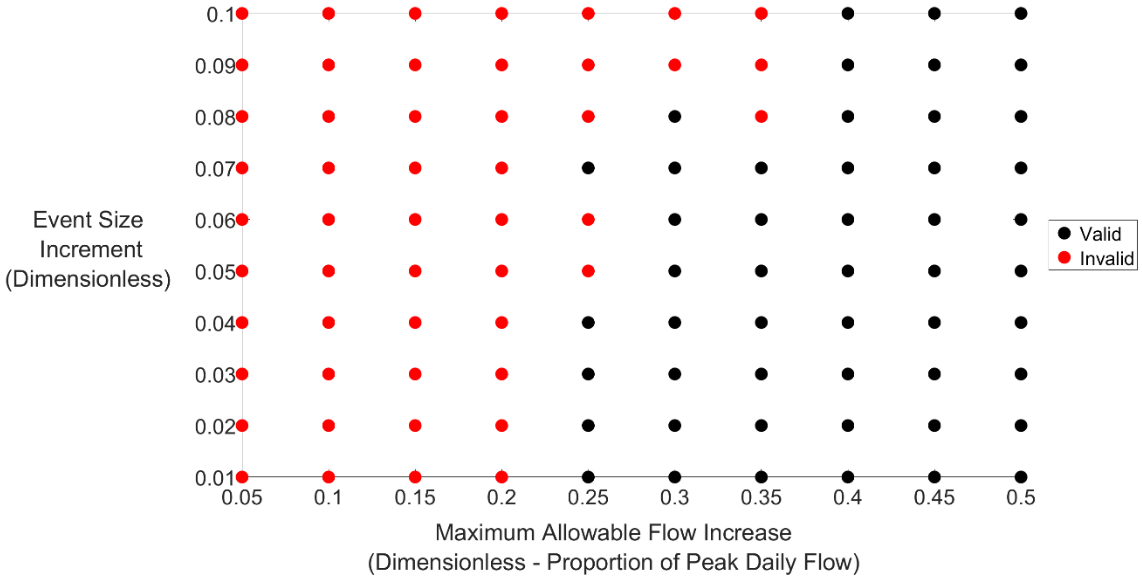


Figure 5-21 *Validity of various combinations of leak/burst event size parameters for the dendritic verification hydraulic model*

The first important observation is that for any maximum allowable flow increase below 0.25, which equates to 25% of the peak daily flow, there were no values for the event size increment which produced valid leak/burst event sizes. This is a property specific to the dendritic verification hydraulic model but must be calculated for each DMA to determine whether they are sensitive enough to unreported leak/burst event sizes. By increasing the maximum allowable flow increase, larger values for the event size increment produced valid results which reduced the number of individual leak/burst event sizes to be considered by the leak/burst event grouping technique. However, this means that the maximum leak/burst can be too large and would likely be localised by other means including customer contacts. The framework of methods was not developed to localise these event sizes and, therefore, increasing the maximum allowable flow increase will violate this constraint. Due to the high value of the maximum allowable flow increase (25% increase in flow) when compared to the 10% value specified in section 4.4.2.3.6 the dendritic verification hydraulic model would not be a suitable DMA for deploying pressure sensors within the context of the framework of methods. This was caused by its synthetic nature because most of the demand was removed from the model during its construction. Therefore, the remaining pipes, which were taken from the inlet of a real DMA hydraulic model, were very large in relation to the demand and the changes in pressure were low as a result. This meant that the dendritic verification hydraulic model was not very sensitive to the maximum leak/burst event sizes.

5.4.6. Case study 5.9: Leak/Burst event grouping procedure

Case study 5.9 provides the results for the application of the leak/burst event grouping procedure to all three of the verification hydraulic models. The main purpose of the leak/burst event grouping technique is to allow multiple leak/burst event sizes to be considered by the sensor placement technique whilst keeping the number of duplicate leak/burst events which are considered to a minimum. In Table 5-5 the number of leak/burst events which would be considered if the leak/burst grouping procedure were not used is given for each of the three hydraulic models in the fourth column. This was calculated as the number of leak/burst event sizes (second column from the left) multiplied by the number of valid leak/burst event locations (third column from the left). The number of leak/burst event scenarios which was produced after the leak/burst event grouping procedure was applied is given. The reduction in the number of leak/burst event scenarios for each of the three hydraulic models was calculated and is provided in Table 5-5.

For each of the three verification hydraulic models considered in Table 5-6 a reduction in the number of leak/burst events considered by the sensor placement technique greater than 80% was achieved. For the dendritic and looped case study networks, which are both smaller and simpler than DMA 123-09, the reduction in the number of leak/burst events was even higher with an approximately 96% reduction. Further to this the number of leak/burst event scenarios remaining after the leak/burst grouping procedure was applied was lower than the case where only a single leak/burst event scenario was applied at all event locations. In the case of the dendritic network if only one leak/burst event size was modelled at all leak/burst event locations then 109 leak/burst event scenarios would be considered. The number of grouped leak/burst event scenarios was 60 which is significantly lower than the 109 leak/burst events when only a single leak/burst event size was considered without the leak/burst event grouping procedure.

Table 5-5 *Reduction in leak/burst event scenarios considered for the three verification hydraulic models*

Case Study Network	Number of Event Sizes	Number of Event Locations	Number of Scenarios (Ungrouped)	Number of Scenarios (Grouped)	% Reduction in Number of Scenarios
Dendritic	18	109	1962	60	96.9
Looped	18	109	1962	75	96.2
DMA 12309	19	425	8075	1486	81.6

Another potential problem with the leak/burst event grouping procedure was related to differences which can occur between leak/burst events in the same group. This problem would lead to leak/burst events which are grouped together to have significantly different search areas and can be caused by the accuracy of the pressure instruments being too large. In effect, the pressure sensor instrument accuracy may allow two events to be grouped together even though the distribution of the measured normalised changes in pressure could point towards the leak/burst event being in a completely different section of the DMA hydraulic model. To test whether this was the case an investigation into the leak/burst event groupings was conducted to determine how similar the search areas were for leak/burst event scenarios which were grouped together.

The results provided in this section pertain to DMA 123-09 as it is the most complex of the verification hydraulic models and provides the most robust test. In order to determine the group of leak/burst events presented here the maximum and minimum leak/burst event sizes were determined and the leak/burst event grouping procedure was applied to the resulting set of sensitivity matrices to determine the groups of leak/burst events for each leak/burst event size. An additional step performed for the purposes of this case study which is not usually required was to extract several groups of leak/burst events so that the examination of the consistency of the search areas could be completed. From the complete set of grouped leak/burst event scenarios some groups were extracted and the search areas for each event in those groups was determined using different optimal sensor configurations. The optimal sensor configurations with 10 sensors were used to determine the search areas for each of the groups of leak/burst events. In Figures 5-22 and 5-23 the search areas for all members of two groups of leak/burst events are shown. The groups of leak/burst events were selected as they contained the most leak/burst events of any group from each of the determined leak/burst event sizes. This was so that the leak/burst locations in the group varied as much as possible to test whether maximising the distance between two leak/burst events in the same group would tend to cause more variation in the search area.

In Figure 5-22 the search areas which were determined using 10 optimal sensors are shown for the one group which contained the most leak/burst events of any group. In total there were four leak/burst events in the group and the search area was identical for each of the leak/burst events. The events in the group were located in a small area spanning only several hundred meters between the two most distant events which makes the similarity between the search areas less surprising. It is most likely that leak/burst events which are close to each other will be grouped together which increases the likelihood of producing similar search areas. A second group of four leak/burst events is shown in Figure 5-23. In this case even though the search areas were not identical they were very similar. The reason that the search areas were not identical is because they events are spread over a greater area than those in group 37. Even in this case the difference between the search areas is very small (only a few nodes and pipes differences) and given that this is the joint largest group the remaining group with fewer events are likely to be identical or very similar as well. This demonstrates that the leak/burst event grouping procedure is removing events which are duplicates in terms of the search areas produced using the optimal sensor configurations as well as the measured changes in pressure which were used to group them.

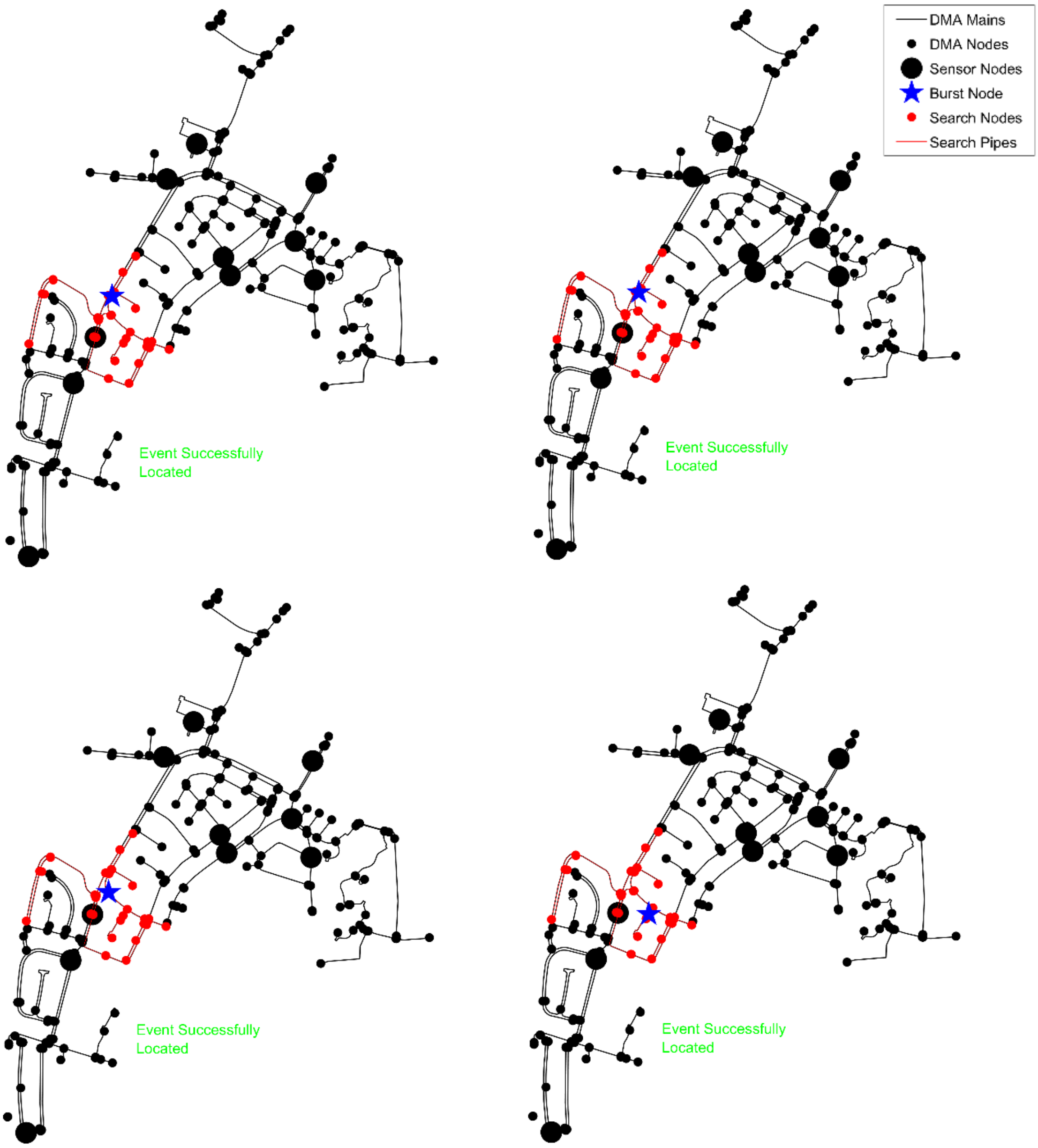


Figure 5-22 *Search Areas for the four leak/burst event locations in group number 37*

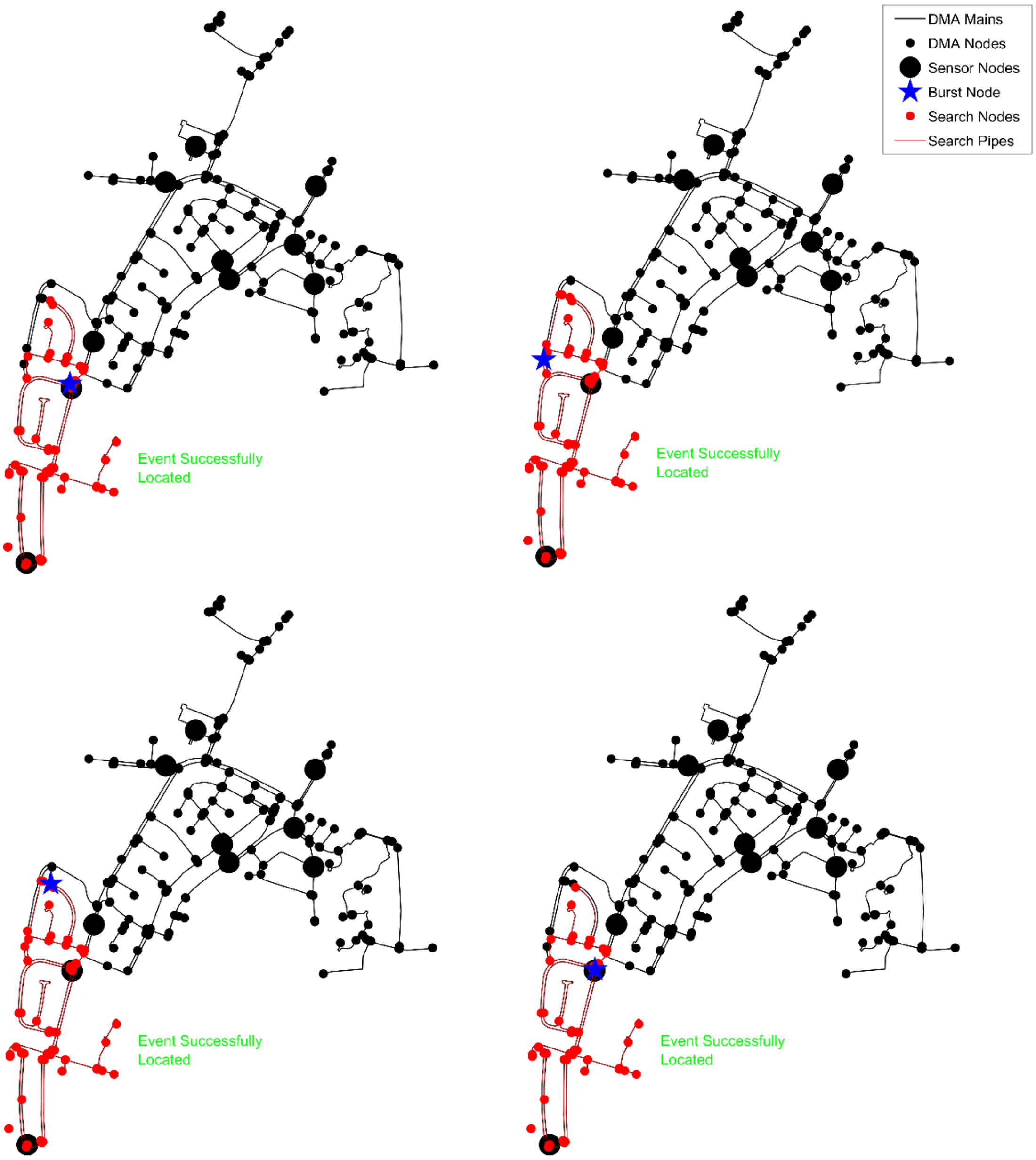


Figure 5-23 *Search Areas for the four leak/burst event locations in group number 99*

In addition to the consistency of the search areas a demonstration of how the leak/burst events are grouped is presented. The rationale behind the leak/burst event grouping technique is that leak/burst events are grouped together based upon the measured changes in pressure at all nodes throughout the hydraulic model. This process does not guarantee how the leak/burst events will be grouped together in terms of their location. The grouping of leak/burst events will depend upon the magnitude of leak/burst events which are modelled as well as the differential sensitivity to leak/burst events throughout the DMA being considered. The grouping of leak/burst events for multiple different sizes of leak/burst events for the dendritic case study network is shown in Figure 5-24. Three different sizes of leak/burst event, corresponding to different emitter coefficients covering the range of event sizes determined by the automatic procedure for determining the maximum and minimum leak/burst event sizes, are demonstrated. The leak/burst event groupings for emitter coefficients of 0.4, 0.13 and 0.21 are shown in Figure 5-24. below. In each case, the leak/burst event locations are grouped according to the colour of the corresponding circles.

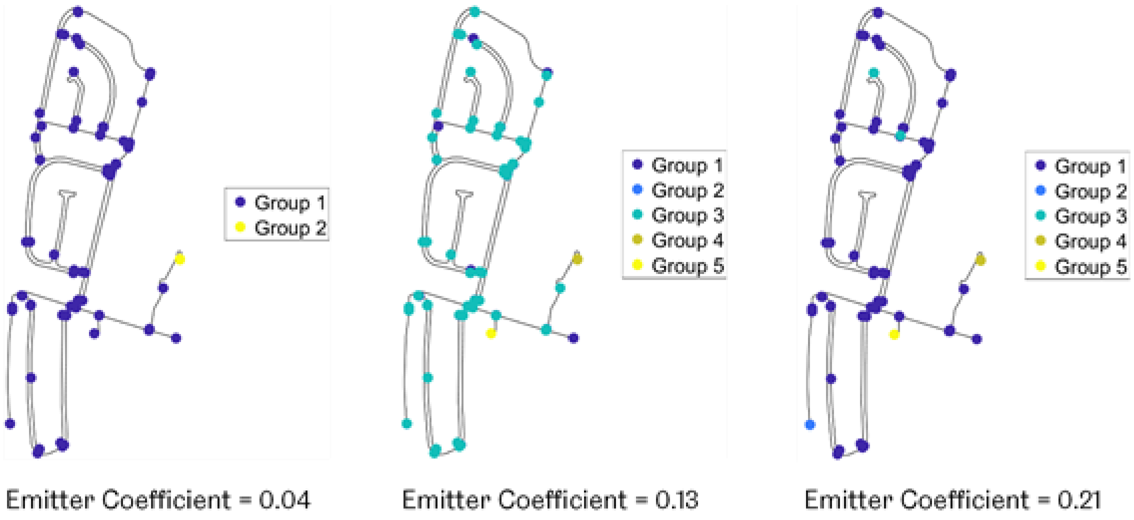


Figure 5-24 *Leak/burst event groupings for three leak/burst event sizes modelled in the dendritic case study network*

Starting with an emitter coefficient of 0.04, on the left of Figure 5-24, only 2 groups of events were determined. The largest group, denoted by the dark blue circles, comprises every leak/burst event location aside from one. The other group, denoted by a yellow circle, is comprised of a single leak/burst event located at the end of pipe leg. For the emitter coefficient of 0.13, in the centre of Figure 5-24, a similar trend was seen except that 5 groups were produced instead of 2. Each of the additional groups contained a single leak/burst event. The emitter coefficient of 0.21 produced 5 groups of leak/burst events. Of the five groups, four were identical to those produced when an emitter coefficient of 0.13 was used. However, one group of events was located in a different part of the DMA. This seems counterintuitive as it seems that additional groups should be produced as the leak/burst event sizes are increased rather than moving one of the groups. As the leak/burst event sizes change the relative changes in pressure throughout the DMA will not be the same for each leak/burst event locations. This means that two leak/burst events that could not be grouped together for smaller leak/burst event sizes may be grouped together as the leak/burst event size increases. Further increases in the leak/burst event size may render the two events too dissimilar to be grouped together again.

5.5. Chapter Summary

In this chapter a number of case studies were presented to demonstrate some of the main performance characteristics of the framework of methods which were described in chapter 3. In relation to the leak/burst localisation technique the most important result was that both the baseline and the normalised leak/burst localisation technique were able to successfully localise real leak/burst events, in the form of engineered events, carried out in a real DMA. The normalised leak/burst localisation technique performed better and was able to localise 8 of the 11 engineered events in DMA 123-09. The average search area for those events which were successfully localised was 25% of the DMA. The localisation case study was performed using a non-optimal sensor configuration with 7 pressure sensors in a DMA containing 1664 properties.

A comparison of two different distance metrics used in combination with the leak/burst localisation technique demonstrated the benefit of using the spatially constrained distance metric which used pipe length instead of Euclidean distance. The main benefit was that pipes which were located close together but were not directly connected to each other can be differentiated by the spatially constrained distance function but not the Euclidean distance function. The result was that the size of the search area was smallest when the spatially constrained distance function was used and the search area was almost 50% smaller than when the Euclidean distance was used.

The first investigation of the sensor placement technique was to explore how the values of both leak/burst localisation parameters affected the performance in terms of the average size of the search area produced and the number of leak/burst events for which multiple search areas were determined. The variation in the average size of the search area was low for small variations in either of the parameters but the number of leak/burst events with multiple search areas behaved much more erratically. The conclusion from this was that to find values of both parameters which led to good performance for both performance metrics using a heuristic approach or "rule of thumb" would not be suitable. Therefore, automatically determining the best value of both parameters in parallel with the sensor configurations has been adopted. The performance of the automatic approach for determining the values of both parameters showed that in most cases the best values of parameters could be determined in only a fraction of the time and without the need to run the sensor placement multiple times. The other advantage of this is that no matter how many values are allowed for either parameter only one run of the sensor placement technique will enable the best (or near-best) values to be determined.

Two additional facets of the sensor placement objective function were explored which were the metric for measuring the size of the search area and the penalty for each time multiple search areas were determined for a leak/burst event. Both of these changes to the objective function were made to address practical problems which could be faced when localising real leak/burst events rather than improving the performance when localising modelled leak/burst events. The penalty for producing multiple search areas produced a significant reduction in the number of leak/burst events with multiple search areas, as intended, but there was a slight increase in the resulting average size of the search area due to the increase in the level of constraint of the problem. By measuring the size of the search area in terms of the length of pipes in the search area as opposed to the count of hydraulic model junctions enables the sensor placement objective function to better match the reality of leak/burst localisation for a field team searching in a DMA. In addition, this step also reduced reliance on the distribution of junctions in a hydraulic model which is arbitrary in relation to the real WDS which the hydraulic model represents. Measuring the search area in this way did not significantly impact the sensor placement performance either positively or negatively in terms of the performance metrics.

Two key parts of the sensor placement technique related to determining the size of leak/burst events and subsequently grouping them together were demonstrated. The automatic determination of leak/burst events makes use of the accuracy of the pressure instruments which are being placed to determine an appropriate range of leak/burst event sizes for each leak/burst event location. This process obviates the need to manually determine the leak/burst event sizes prior to running the sensor placement technique which can be a costly investigative step that is used by several other sensor placement techniques. Two parameters, which are the event size increment and the maximum allowable increase in flow, need to be specified prior to determining the leak/burst event sizes automatically. This allows control over the size of leak/burst events which are considered so that the sensor placement technique can be tailored to the sizes of events which are of concern. The event size increment will control the accuracy with which the maximum and minimum leak/burst event sizes are determined but comes at the cost of the number of leak/burst event sizes which are considered between the maximum and minimum leak/burst event sizes. The two parameters also interact such that improper selection of either will lead to invalid results being produced. For example, if the maximum allowable increase in flow is too small then the maximum leak/burst event sizes can be smaller than the minimum leak/burst event sizes and no valid leak/burst event sizes can be considered. Similarly, if too large an event size increment is selected then the maximum and minimum leak/burst event sizes will not be determined accurately and a valid range of event sizes cannot be formed between the two.

The leak/burst event grouping technique was applied to three different hydraulic models and demonstrated a reduction in the number of leak/burst events which must be considered by the sensor placement technique. The reduction was at least 80% and was as high as 96% for the two simpler hydraulic models considered. The advantage of this approach is that multiple leak/burst event sizes can be considered by the sensor placement technique whilst also considering fewer leak/burst events. This is achieved by removing a number of duplicate leak/burst events in terms of the similarity between the changes in pressure that they cause. Further to this, groups of leak/burst events were localised and the similarity between the search areas was checked. Even when there were several hundred metres of separation between the leak/burst event locations the search areas were almost identical. This was demonstrated for two of the groups with the most leak/burst events in them and with the largest amount of physical separation.

Chapter 6 - Case Studies for Validation of the Framework of Methods

6.1. Introduction

In Chapter 5, several case studies were presented which demonstrated key developmental steps in the framework of methods to verify its performance. A combination of two simple hydraulic models, derived from a hydraulic model of a real DMA, and a complete real DMA hydraulic model were used for each of these case studies to evaluate the performance of the sensor placement technique. Eleven engineered events conducted in a real DMA in United Utilities WDS were used to demonstrate the performance of the leak/burst localisation technique, however, the sensor placement technique was not used to place the sensors or evaluate the performance of the sensor configurations. Therefore, further testing of the framework of methods on previously unseen DMAs was completed, and is described in this chapter, to validate the framework of methods. The combined performance of the sensor placement technique and the leak/burst localisation technique was determined by applying them both to the same DMAs, in the manner that they were designed to be generically used. A case study involving two previously unseen (to the framework of methods) DMAs was completed and is described in this chapter. As with the case studies provided in Chapter 5, related to the development of the framework of methods, data from engineered events and industry standard hydraulic models were used. Due to limitations resulting from the field work a constrained version of the sensor placement technique has been used to validate the framework of methods.

There were several objectives for the validation case study presented in this chapter which are listed below.

1. To deploy additional pressure sensors in two real DMAs and utilise a constrained version of the sensor placement technique to evaluate the

performance of the possible sensor configurations which could be formed from the deployed sensor locations. The constraint was applied to the sensor placement technique to reflect the pre-selected non-optimal sensor locations which were deployed prior to the engineered events.

2. To evaluate the combined ability of the sensor placement and leak/burst localisation techniques to successfully and accurately localise real leak/burst events using engineered events performed in two real DMAs. The best configurations of sensors identified by the constrained sensor placement technique, with varying numbers of sensors, were used in combination with the optimal combination of parameters.

3. To determine how the number of deployed sensors used to localise each engineered event affected the number of engineered events which were correctly localised and the average size of the search areas produced.

4. To identify leak/burst event characteristics which influenced the leak/burst localisation performance.

6.2. Validation Approach

6.2.1. Overview

The key test for all leak/burst localisation techniques is whether they can successfully and accurately localise leak/burst events occurring in real WDSs. This is particularly relevant for the framework of methods presented in this thesis as it does not rely on the use of a highly calibrated hydraulic model and therefore any localisation results generated during the sensor placement step cannot be used directly to demonstrate the leak/burst localisation performance. To allow the leak/burst localisation performance of the framework of methods to be demonstrated a series of real-world engineered events were performed in two validation DMAs. Engineered events are used as a proxy for real leak/burst events which will eventually be localised by the framework of methods. Engineered events are simulated leak/burst events introduced into a WDS by opening fire hydrants in the field and are used because they allow the run time, location and size of the event to be controlled. This is an advantage of engineered events over using real (arising as a result of no operator intervention) leak/burst events.

6.2.2. Selection of Validation DMAs

Two DMAs were selected to validate the framework of methods which was described in Chapter 3 of this thesis. In this chapter, the two DMAs used to validate the framework of methods will be referred to as the validation DMAs. The validation DMAs were selected from one water supply zone (WSZ), referred to as WSZ 304, in United Utilities WDS. Figure 6-1 shows how the validation DMAs are connected to Coppice service reservoir.

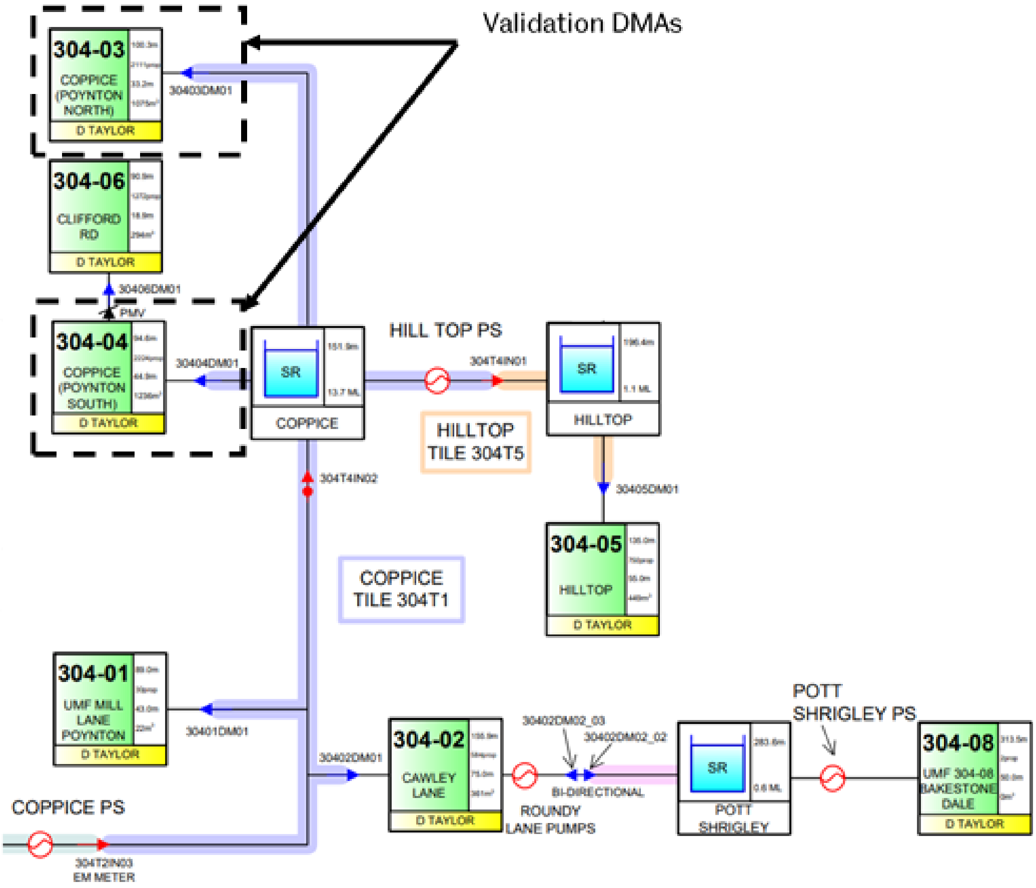


__Figure 6-1__ Overview of WSZ 304 from which contains the validation DMAs. The validation DMAs have been highlighted

The validation DMAs are located in the town of Poynton, Cheshire which is approximately 10 miles outside of Manchester in the Northwest of the United Kingdom. Poynton has a population of approximately 15,000. The validation DMAs, which are called DMA 304-03 and DMA 304-04 are fed from Coppice service reservoir which is located approximately 1km from the inlet of DMA 304-04. There were several reasons for selecting the validation DMAs from WSZ 304. Firstly, all of the DMAs are fed from only one service reservoir. This means that the flows and pressures are more likely to follow a regular pattern as opposed to those DMAs which can be fed from multiple sources. When a DMA can be fed from multiple service reservoirs, which is sometimes referred to as a "push-pull" system, large and abrupt variations in the night-time pressures, which are used by the framework of methods, are possible. The framework of methods relies on a set of "normal" previous days for which the night-time pressures are stable and consistent to determine the changes in pressure which have occurred at each sensor location. Selecting DMAs which are fed from a single service reservoir increases the consistency between the normal days by removing the possibility of significant changes in the behaviour of the WSZ upstream of the validation DMAs. DMA 304-04 is part of a cascading system and DMA 304-06 is connected to Coppice service reservoir via DMA 304-04. It was highlighted by Romano et al. (2011) that cascading DMAs present additional challenges when compared to DMAs which are connected directly to transmission mains without passing through any other DMA. This is because changes in the behaviour of one DMA in a cascading system can impact on the other DMAs.

The validation DMAs are situated in a small town and are urban in nature. This was an important factor in the selection of the validation DMAs as the majority of United Utilities customers are situated in urban DMAs. The framework of methods aims to improve customer service (amongst other things) by reducing the impact that leak/burst events have on customers so focussing on the DMAs which contain the most customers will allow the biggest impact to be seen for all customers across United Utilities entire WDS. This line of reasoning is also applicable to any water company in developed countries whose customers predominantly live in urban centres.

Another pertinent factor was that several other projects to install sensors in WDSs were being conducted in this WSZ. In the interest of trying to minimise the overall disruption to customers it was logical to use the same WSZ for the validation of the framework of methods. This is because there will always be some disruption to customers when works are being conducted in WDSs. For example, even for a relatively simple task such as installing a pressure sensor, a path or road may need to be closed. By using the same WSZ where other works were already being conducted this reduced the number of customers who could potentially be impacted and means that the works could be carried out under the same customer notifications that had been served for the other projects.

6.2.3. Overview of Validation DMAs

The validation DMAs are adjacent to each other and the location of the two validation DMAs, highlighted using blue and red polygons, are shown in Figure 6-2. The feed from Coppice Service reservoir enters DMA 304-03 and DMA 304-04 to the right of Figure 6-2 and the point of connection between DMA 304-04 and DMA 304-06 is labelled. Due to the undulating nature of the local terrain each of the DMAs contain pressure reducing valves (PRVs). DMA 304-03 is split into three discrete pressure areas (DPAs) and the pressure in two of these is controlled using a PRV. The PRVs are used to regulate the pressure in each DPA at a sufficient level as to maintain the minimum required pressure whilst also ensuring that there is not excessive pressure. Excessive pressure causes increased levels of leakage and increases the stress in pipe walls which can cause higher numbers of pipe failures. In the case of DMA 304-03 the majority of the DMA is fed under gravity with no regulation of the pressures. DMA 304-04 contains three DPAs and each are approximately equal in size. DMA 304-06 contains a PRV at the inlet from DMA 304-04. For each validation DMA the pipes in each DPA are shown in Figures 6-3 and 6-4 in the next section.

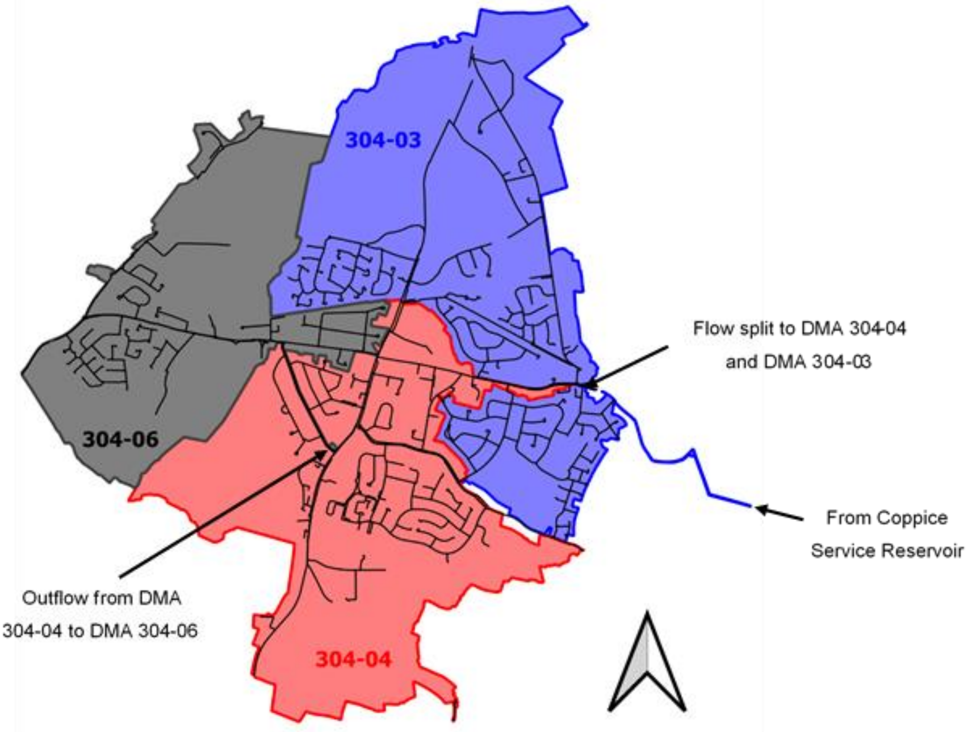


Figure 6-2 *Overview of the two validation DMAs (304-03 and 304-04) and DMA 304-06 which cascades from DMA 304-04*

6.2.4. Selection of Deployed Sensor Locations

In the two validation DMAs, 29 pressure sensors were deployed for various purposes. 12 sensors were deployed in DMA 304-03 and 17 were deployed in DMA 304-04. Two different types of pressure sensors were deployed. In DMA 304-03, Cello pressure sensors supplied by Technolog™ were deployed alongside pressure sensors supplied by Inflowmatix™. For DMA 304-04 only Inflowmatix™ pressure sensors were deployed. All of the pressure sensors collected a pressure measurement every minute. The pressure sensors were deployed in December 2019 and the data was collected until March 2020 which covered the period during which the engineered events were conducted in the validation DMAs. The pressure sensors were being used for multiple projects concurrently, as such the optimal configurations as determined by the optimal sensor placement technique could not be deployed. For each of the validation DMAs the sensors were manually spread evenly throughout the DMAs to ensure that the entire DMA, including all DPAs, was covered in spatial terms. For each of the validation DMAs a map of the deployed sensor locations is given in Figures 6-3 and 6-4 below. The sensor indices given in Figure 6-3 and Figure 6-4 are used to refer to each of the sensors throughout the remainder of this chapter.

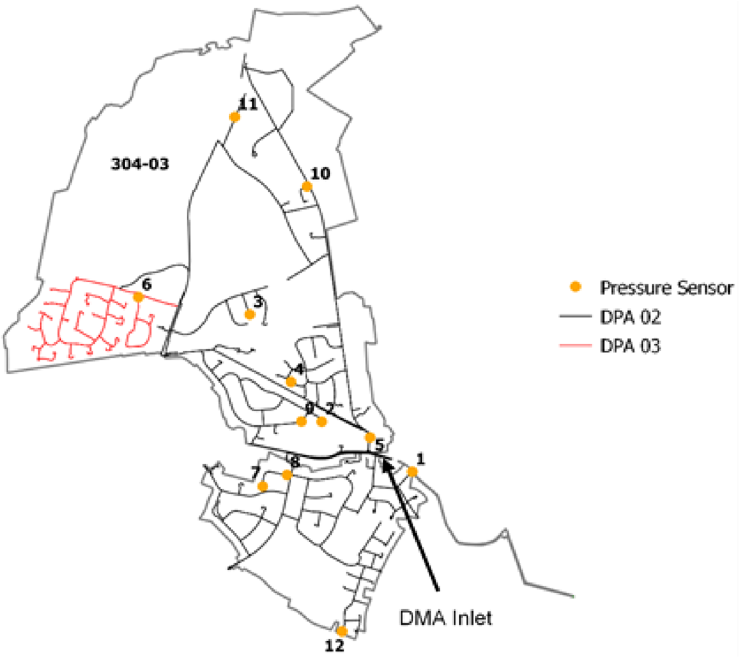


Figure 6-3 *The 12 deployed sensor locations and DPAs for DMA 304-03*

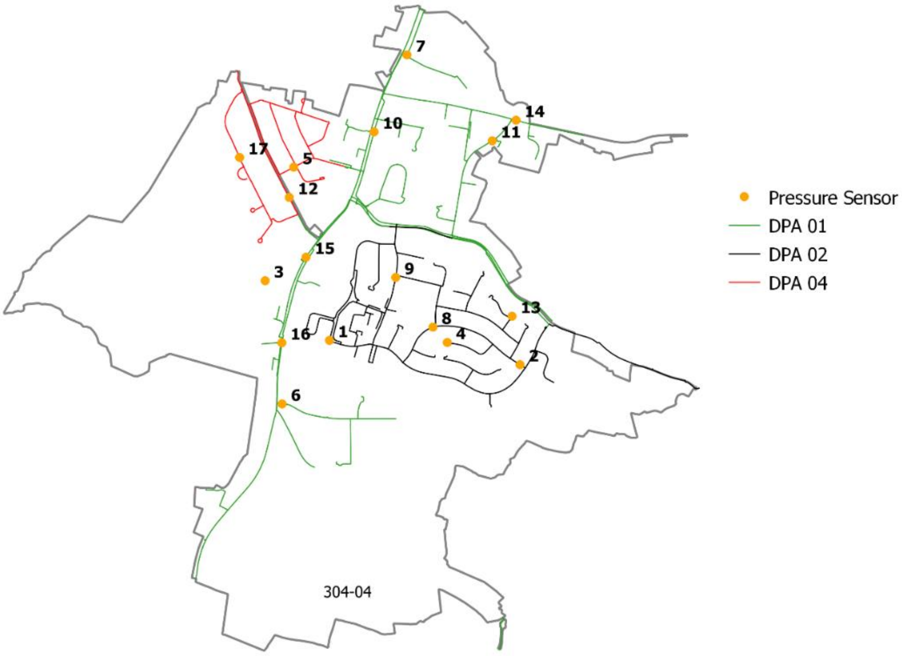


Figure 6-4 *The 17 deployed sensor locations and DPAs for DMA 304-04*

6.2.5. Selection of Engineered Event Locations

In a similar fashion to the deployed sensor locations, the engineered event locations were manually chosen to ensure that at least one engineered event was conducted in each section of the DMA. This is particularly pertinent given that each of the validation DMAs are divided into DPAs and it was desirable to ensure that each of the DPAs contained an event.

The engineered events were conducted during the period of minimum flow which is used by the leak/burst localisation technique. Each event was started before 03:00am and ended after 04:00am. The opening of each hydrant was controlled to achieve an approximate flow rate of 0.6 l/s at the start of the event and the same hydrant opening was maintained for the duration of the engineered events. The size of the events was determined by the engineers conducting the field work however this was approximately half the size of the smallest events determined by the automatic procedure for determining the leak/burst event sizes used by the sensor placement technique. The locations of the engineered events were selected so as not to include any of the hydrants with sensors installed. The potential impact of this is these engineered events would be easily localised and not fully test the performance of the leak/burst localisation technique. This equated to approximately 6% of the average (or 3.5% of the peak) daily flow to the DMA, calculated over a normal week, for DMA 304-03. For DMA 304-04 the same flow rate of 0.6 l/s was used which equated to 4.6% of average (or 3% of the peak) daily flow. The sizes of the engineered events were selected to be sufficiently large to be distinguishable from anomalous but legitimate customer demand which could still occur, even during the night. The sizes of the engineered events were, however, significantly smaller than in other studies which used them to validate leak/burst detection and localisation techniques such as Farley et al. (2010) or Romano et al. (2014). This was true in absolute terms and also proportionally to the DMA inflow. The flow rate was measured by using a flow meter installed in the standpipe connected to the hydrant but was not checked further aside from when setting the position of the hydrant initially. Any flow from the standpipe was diverted to a drain or road gully to prevent flooding or pooling water in the residential areas contained within the DMAs.

The locations of the engineered events for the validation DMAs are shown as a red circle in Figures 6-5 and 6-6 for DMA 304-03 and DMA 304-04, respectively. The indices for each event location are given by the red numbers next to the locations. In DMA 304-03 some event locations were used multiple times on different days such that 8 unique hydrants were opened across the 16 engineered events. In DMA 304-04, 4 different hydrants were used for the 4 engineered events.

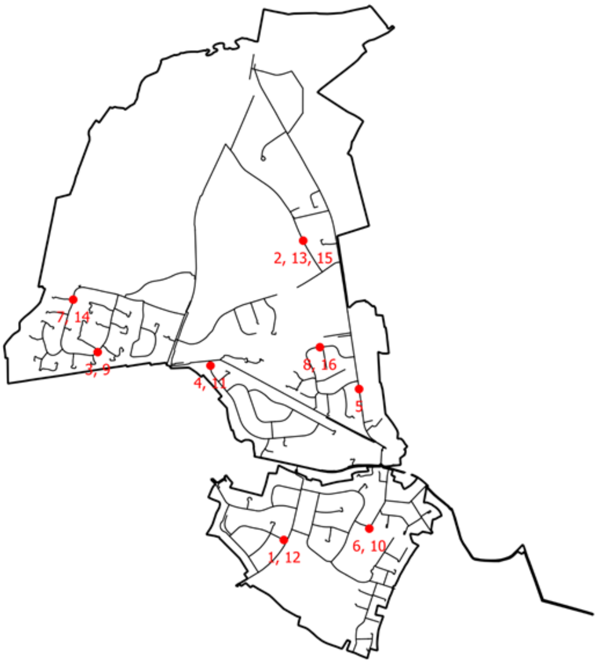


Figure 6-5 *Locations and indices for the engineered events in DMA 304-03*

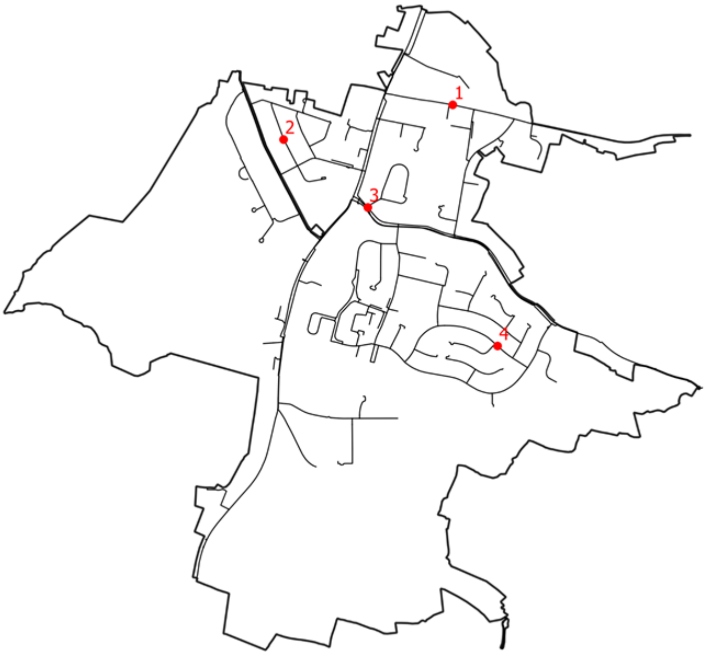


Figure 6-6 *Locations and indices for the engineered events in DMA 304-04*

6.2.6. Evaluation and Selection of Sensor Configurations

As was stated in Chapter 4, due to the combinatorial nature of the sensor placement problem, there are multiple different sensor configurations which can be made from a set of sensor locations. The number of possible combinations with a specific number of sensors is governed by the number of sensors which can be chosen from and the number of sensors which are to be selected from the available locations. The sensor locations given in Figures 6-3 and 6-4 were added as a constraint to the sensor placement technique so that only these locations could be selected from. The sensor placement objective function, using the average size of search areas produced for all modelled leak/burst events, was used to identify the optimal sensor configurations and combinations of parameters. As more than 10 sensors were deployed in both of the validation DMAs the optimal sensor configurations with between 3 and 10 sensors (inclusive) were determined.

In addition to identifying the optimal sensor configurations some baseline sensor configurations were selected at random from the available sensor locations in each validation DMA. Baseline configurations with between 3 and 10 sensors (inclusive) were selected so that a comparison of the performance of the baseline and optimal sensor configurations could be completed for all numbers of sensors considered. For the baseline sensor configurations, the best value of the leak/burst localisation parameters was determined by evaluating all possible combinations. For each baseline sensor configuration, the best combination of parameters was used because selecting the parameters at random would be extremely detrimental to the achievable localisation performance.

6.2.7. Leak/Burst Localisation Performance Comparison

One of the objectives of the case studies provided in this chapter was to demonstrate the leak/burst localisation performance of the optimal sensor configurations, determined by the sensor placement technique, on real leak/burst events. This section will describe the analysis procedure which was used for this purpose.

For both validation DMAs, once the optimal sensor configurations, which were constrained by the deployed sensor locations, and combinations of parameters were determined the leak/burst localisation performance was determined for each engineered event. For each engineered event the optimal sensor configuration and combination of parameters was used to determine the size of the search area, whether the event was correctly localised and whether multiple search areas were determined. The average size of the search area, across all engineered events, was used as the measure of performance for each of the optimal sensor configurations. The same penalties for producing multiple search areas and incorrectly localising leak/burst events used by the sensor placement objective function were applied to the engineered events. This was so that a direct comparison could be made between the leak/burst localisation performance achieved using modelled data and real data collected during engineered events.

To evaluate how well the optimal sensor configurations performed, a baseline level of performance was determined. The baseline level of performance was determined using the baseline sensor configurations and leak/burst localisation parameter combinations which were selected according to the procedure described in section 6.2.6. The purpose behind this was to represent the typical level of leak/burst localisation performance which would be achieved if sensors were deployed at random without using the sensor placement technique.

6.3. Validation Case Study Results

6.3.1. Overview

The results of the application of the framework of methods to each of the validation DMAs are provided in this section. For each validation DMA, the results of the constrained sensor placement technique are provided first followed by the leak/burst localisation results which were derived from the optimal and baseline sensor configurations. For DMA 304-03, because more engineered events were conducted there, further analysis of the leak/burst localisation performance was conducted to provide further insights into the performance of the framework of methods.

6.3.2. DMA 304-03

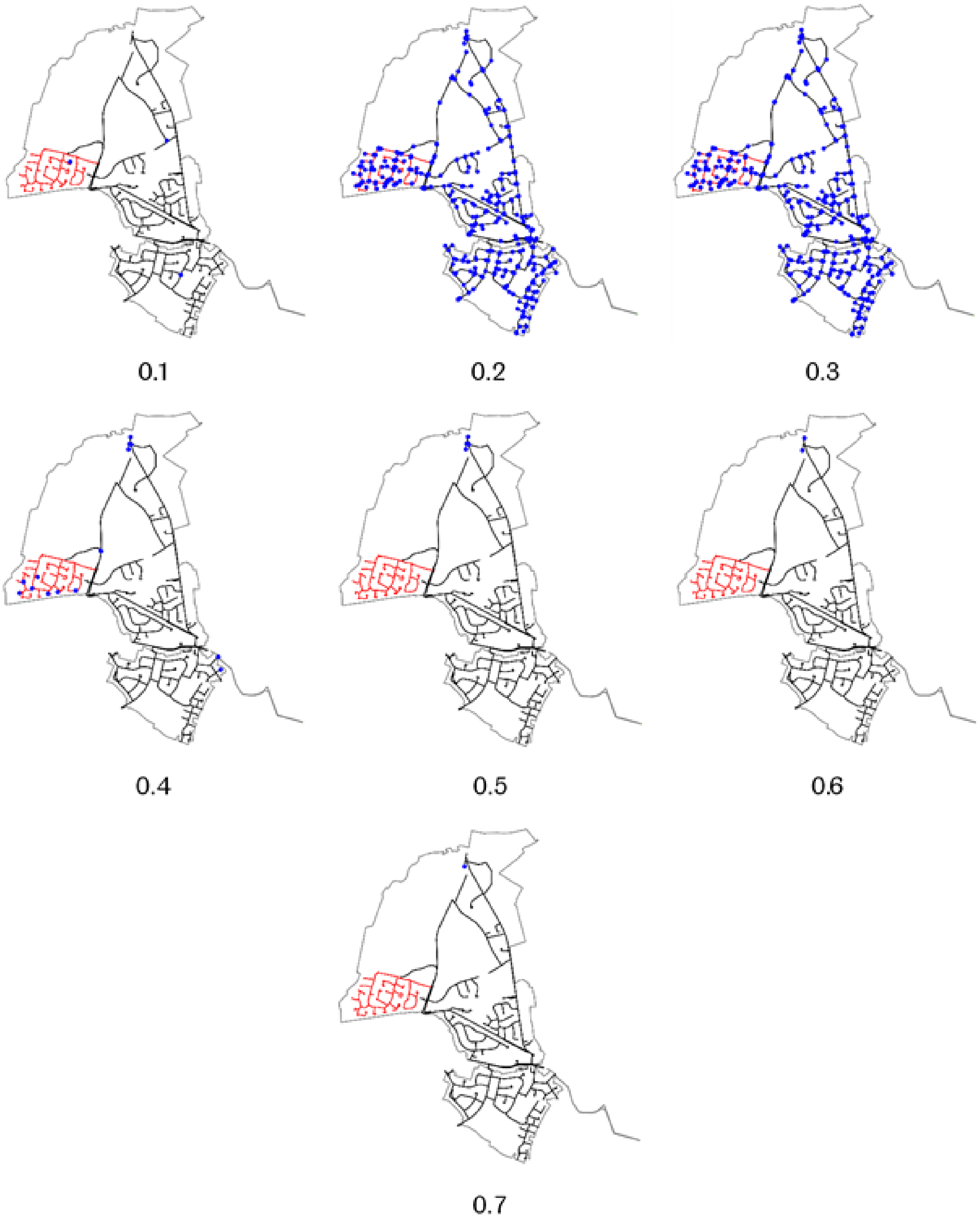
6.3.2.1. Leak/Burst Event Sizes and Leak/Burst Event Grouping

To determine the maximum and minimum leak/burst event sizes an event size increment of 0.1 was specified. The maximum allowable increase in flow was specified as 0.1 which equated to a 10% increase in the flow relative to the maximum daily flow. The automatically determined leak/burst event sizes and the corresponding results from the leak/burst event grouping technique, which was described in section 4.4.2.3.7, are given in Table 6-1 below. For each of the considered leak/burst event sizes, given in the first column, the number of valid leak/burst event groups produced are given in the second column. In this case, the total number of leak/burst event scenarios considered by the sensor placement technique was 934. In the ungrouped case there were 731 leak/burst event locations and combining this with the 7 leak/burst event sizes a total of 5117 leak/burst event scenarios could be considered. This represents a reduction of approximately 82%. Figure 6-7 shows the spatial distribution of the grouped leak/burst event scenarios for each of the leak/burst event sizes given in Table 6-1. Each of the considered leak/burst event scenarios are shown as a blue circle.

Table 6-1 *Leak/burst event sizes and leak/burst event grouping results for DMA 304-03*

Leak/burst event size	**Number of valid leak/burst event groups**
0.1	3
0.2	437
0.3	469
0.4	18
0.5	4
0.6	2
0.7	1
Total	**934**

The majority of the leak/burst events included in the sensitivity matrix considered by the sensor placement technique were for leak/burst event sizes (emitter coefficients) of 0.2 and 0.3. The reason why the number of valid leak/burst event groups is highest for these leak/burst event sizes was caused by two things. Firstly, the total number of groups was higher because there was more variation between the sensitivities for individual leak/burst events. However, checking the validity of the groups of leak/burst events has the effect of removing groups and the number of invalid groups increased as the leak/burst event size did because the larger leak/burst event sizes did not fall within the range of valid leak/burst events for most groups. This is logical as the smallest leak/burst event sizes will only cause a measurable change in pressure at the most sensitive leak/burst event locations. Similarly, for the largest leak/burst event sizes, only the least sensitive leak/burst event locations will not violate the maximum increase in flow specified.



0.1 0.2 0.3

0.4 0.5 0.6

0.7

Figure 6-7 *Grouped leak/burst event locations for each of the emitter coefficients determined by the automatic leak/burst event size determination procedure for DMA 304-03*

6.3.2.2. Optimal and Baseline Sensor Configurations and Parameters

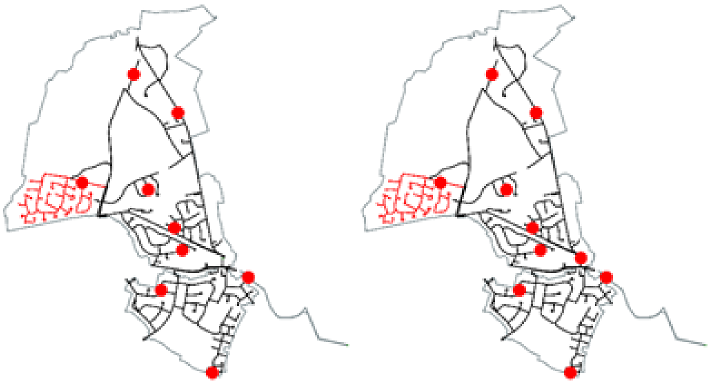
In this section one table and two figures are included which relate to the optimal and baseline sensor configurations and leak/burst localisation parameters. The optimal sensor configurations and parameters were determined using the constrained sensor placement technique. The optimal sensor configurations were selected from the available deployed sensor locations, of which there were 12 for DMA 304-03. The optimal parameters were determined using the constrained version of the sensor placement technique. For each number of sensors, the baseline sensor configurations were selected by randomly sampling from the available deployed sensor locations, with no repeats. For each baseline sensor configuration, the best combination of parameters was determined by evaluating all combinations of parameters and selecting the best. In Table 6-2, the best combinations of search area threshold and interpolation exponent for each constrained optimal and baseline sensor configuration are given. In Figure 6-8, the constrained optimal sensor configurations are denoted by red circles and in Figure 6-9 the baseline sensor configurations are denoted by green circles.

Table 6-2 *Leak/burst localisation parameters determined for the baseline and optimal sensor configurations for DMA 304-03*

Number of Sensors	**Baseline**		**Optimal**	
	Search area threshold	**Interpolation exponent**	**Search area threshold**	**Interpolation exponent**
3	0.9	6	0.95	15
4	0.85	24	0.8	8
5	0.95	8	0.9	15
6	0.8	28	0.95	15
7	0.75	13	0.95	8
8	0.65	9	0.95	6
9	0.9	18	0.95	10
10	0.95	30	0.95	15

3 Sensors 4 Sensors 5 Sensors

6 Sensors 7 Sensors 8 Sensors



9 Sensors 10 Sensors

Figure 6-8 *Constrained optimal sensor configurations (red circles) with varying numbers of sensors for DMA 304-03*

3 Sensors 4 Sensors 5 Sensors

6 Sensors 7 Sensors 8 Sensors

9 Sensors 10 Sensors

Figure 6-9 *Baseline sensor configurations (green circles) with varying numbers of sensors for DMA 304-03*

The constraint placed on the sensor placement technique meant that as the number of sensors increased so did the similarity between the constrained optimal and baseline sensor configurations. This is because the maximum number of sensors considered by the constrained sensor placement techniques was similar to the number of deployed sensors. The constrained optimal and baseline sensor configurations with 10 sensors must have, at least, 8 sensor locations in common because there are only two additional sensors which could differ. For the sensor configurations with 3 sensors the likelihood of having any sensor locations in common between the configurations was significantly lower than for the case with 10 sensors.

6.3.2.3. Comparison of the constrained optimal and baseline Sensor placement performance

The Pareto front for the constrained optimal sensor configurations and parameters and the performance achieved for the baseline sensor configurations and parameters are shown in Figure 6-10 for DMA 304-03. For all numbers of sensors, the constrained optimal sensor configurations outperformed the baseline sensor configurations, as expected. The difference in performance between the optimal and baseline configurations and parameters ranged from 18.1%, for the cases with 3 sensors, and 1.9% for the case with 10 sensors. Due to the random nature of the baseline sensor configurations the performance did not improve with each added sensor although the general trend was that the performance tended towards the optimal level of performance as the number of sensors increased. This occurred because the number of deployed sensors was only slightly higher than the maximum of sensors considered for this DMA. For the case of selecting 10 sensors from the 12 deployed sensors there were 66 possible sensor configurations which could be selected whereas for the case with 3 sensors there were 220 possible configurations which could be selected. It was logical, therefore, that the difference diminished as the number of sensors considered tended towards to available number of deployed sensors.

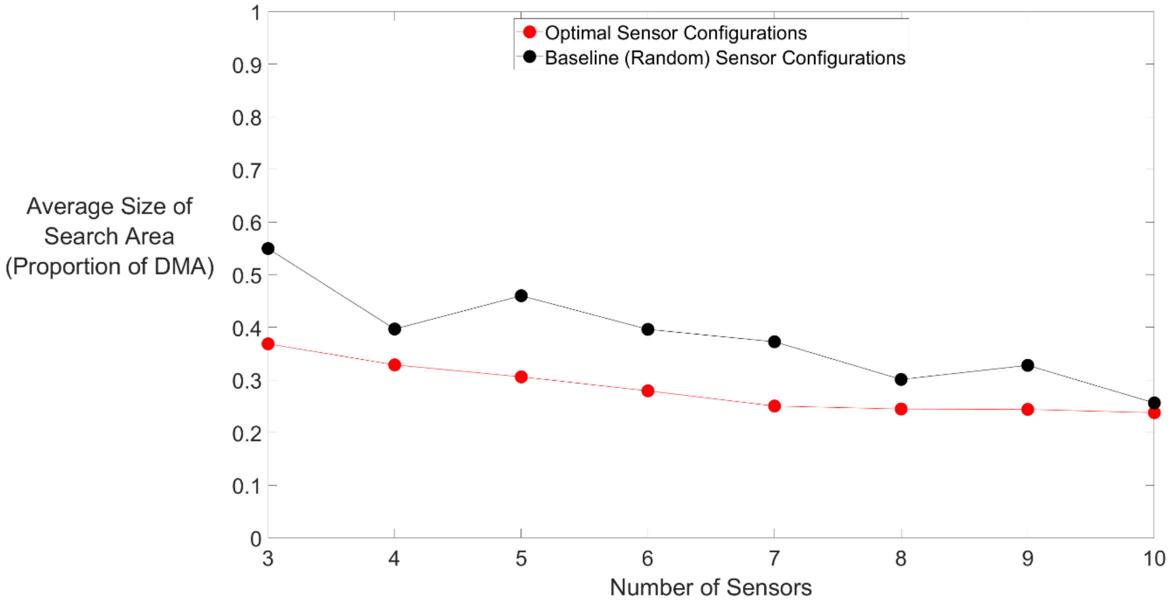


Figure 6-10 *Comparison of the performance for the constrained optimal and baseline sensor configurations for DMA 304-03*

6.3.2.4. Comparison of the leak/burst localisation performance using the engineered event data

The performance of the leak/burst localisation technique, when applied to the each of the engineered events in DMA 304-03, is given in Table 6-3 for the constrained optimal sensor configurations and in Table 6-4 for the baseline sensor configurations. For each engineered event, the size of the search area and whether the event was correctly localised are shown for each optimal sensor configuration. Correctly localised events are denoted by bold text and an asterisk denotes events for which multiple search areas were produced. At the bottom of each table of results the equivalent value for the objective function, accounting for both objective function penalties, is given so that results obtained using the modelled data and the real data can be compared. In addition to comparing the sensor placement results (considering the grouped leak/burst event scenarios) with the leak/burst localisation results, a comparison between the modelled results using the same size and locations as the engineered events was conducted. This is because there will always be some discrepancy between the results achieved by the sensor placement technique and the leak/burst localisation technique because they were not analysing the same set of leak/burst event scenarios. This effect was also magnified because the engineered events were performed in the same locations on multiple days which would favour sensor configurations which perform well for the engineered event locations regardless of whether they are optimal with respect to all possible leak/burst locations considered by the sensor placement technique.

Considering all optimal sensor configurations, the average size of the search area produced for all engineered events was approximately 21% of the DMA. Of the 128 total combinations of optimal sensor configuration and engineered event 49 were correctly localised. Comparing this with the aggregated performance of the baseline sensor configurations for which the average search area was 26% and 43 combinations of baseline sensor configuration and engineered event were correctly localised. On average the optimal sensor configurations produced smaller search areas and correctly localised more events than the baseline sensor configurations. The number of leak/burst events with multiple search areas was 21 across all baseline sensor configurations compared to 28 across all of the optimal sensor configurations, however.

Table 6-3 *Results for the optimal sensor configurations for the 16 engineered events in DMA 304-03. Events in bold were correctly localised and events marked with an asterisk (*) produced multiple search areas*

Engineered Event Number	**Number of Optimal Sensors**							
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	**38.7**	33.2	**36.2**	**61.2**	**61.2**	**61.1**	**61.3**	**61.3**
2	30.7	**29.5**	30.7	**9.5***	**9.5***	**15.4***	**13.5***	0.8
3	23.4	20.8	23.4	2.4	2.4	9.9*	10.1*	14.4
4	39.4	**22.1**	**35.5**	6.2	6.1	14.3	15.6	18.4
5	30.7	**65.9**	25.1	**20.9**	**20.9**	**17.6**	**18.7**	**11.9**
6	28.6	28.4	25.1	2.4	2.4	2.4	2.4	2.4
7	**28.6**	28.9	**25.1**	8.6*	8.6*	5.6	5.8	5.8
8	**23.4**	31.5	**18.5**	**23.3***	**23.3***	**18.0***	**20.1***	**14.2***
9	**28.3**	29.5	**25.1**	2.4	2.4	2.4	2.4	2.4
10	28.3	28.4	25.1	2.4	2.4	2.4	2.4	2.4
11	**28.6**	**27.1**	25.1	8.6*	8.6*	5.6	5.8	5.8
12	28.3	21.4*	**35.5**	20.9	20.9	**39.7**	**39.9**	**31.1***
13	30.7	**29.5**	25.0	2.4	2.4	2.4	2.4	2.4
14	**30.7**	21.7	**65.0**	8.5*	6.1	8.1*	8.2*	8.2*
15	30.7	**65.9**	**37.9**	29.4*	29.4*	23.3*	26.9*	20.1*
16	**54.3**	29.5	**43.6**	**28.9**	**28.6**	**23.4**	**23.5**	**20.4**
Average Search Area	33.2	40.0	35.8	28.8	28.7	29.3	29.5	27.8
Correct Events	7	6	9	5	5	6	6	5
Events with multiple search areas	0	1	1	6	5	5	6	4
Equivalent objective function value	70.8	77.5	63.9	88.2	88.2	83.9	84.0	87.1

Table 6-4 *Results for the baseline sensor configurations for the 16 engineered events in DMA 304-03. Events in bold were correctly localised and events marked with an asterisk (*) produced multiple search areas*

Engineered Event Number	**Number of Baseline Sensors**							
	3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	**49.1**	**59.4**	**49.1**	**46**	**52.3**	**71.6**	**29.2**	**63.3**
2	48.9	**10.3**	2.4	**52.5***	38.3	25.2	29.9	24.9
3	49.2	**59.3**	2.4	**45.1***	38.3	25.2	29.9	24.9
4	48.9	**33.5***	17.3	**22.2**	**16.2**	6.1	6.1	6.1
5	49.2	**46.4**	25.9	14.2	**57.5**	**25.7**	**20.9**	13.8
6	2.4	27.3	2.4	27.4	11.1	41.8*	26.2	2.6
7	49.2	33.5	6.7*	15.1	15.8	32.0	7.9	6.7*
8	32.5	25.5	2.4	**41.8**	16.0	**29.3***	25.5	**16.5***
9	44.6	**54.1**	2.4	**33.2***	16.0	28.1	**25.5**	2.6
10	2.4	25.5	2.4	27.4	15.1	20.3	25.5	2.6
11	51.6*	**33.5**	6.7*	**10.2**	**15.1**	**30.0**	**7.0**	6.7*
12	51.6*	33.5	**36.5**	14.1	**34.9**	**49.1**	6.1	20.0*
13	2.4	**35.2**	2.4	27.4	**28.1**	**28.1**	27.3	2.6
14	49.2	**59.3**	6.3*	**35.1***	12.9	22.4	7.0	6.6*
15	49.2	33.5	8.1*	58.4	30.1*	**61.6**	8.0	20.4*
16	12.9*	10.0	2.4	**22.5**	22.9	**54.2**	8.7	**22.2**
Average search area	50.4	43.4	42.8	34.3	40.0	43.7	20.7	34
Correct events	2	9	2	9	6	8	4	3
Events with multiple search areas	3	1	4	4	1	2	0	6
Equivalent objective function value	96.8	68.2	92.8	77.7	75.3	76.3	80.2	92.8

6.3.2.5. Comparison of modelled and real changes in pressure

In this section further investigation of one particular engineered event which was successfully localised was used to demonstrate some of the characteristics of the leak/burst localisation technique. The engineered event selected was number 8 (which was correctly localised by seven of the eight optimal sensor configurations). The engineered event data is given in Figure 6-11 below. The raw changes in pressure, determined using the hydraulic model and for the engineered event, are plotted for each sensor. The changes in pressure are given for all 12 pressure sensors which were deployed for the engineered events.

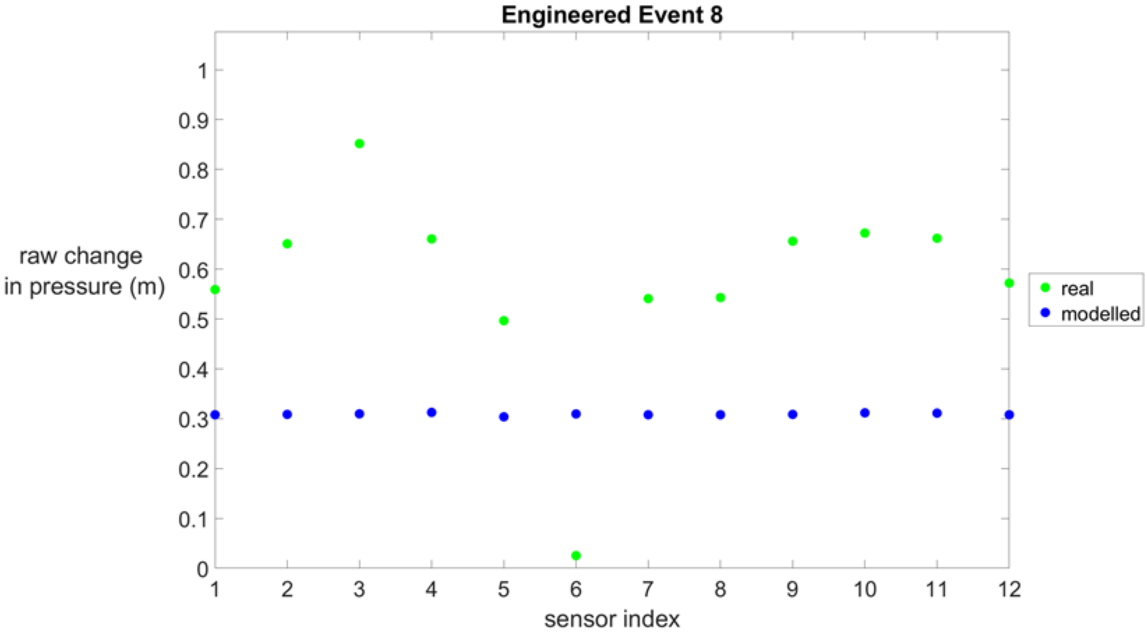


Figure 6-11 *Raw changes in pressure for engineered event 8 determined using the hydraulic model (blue circles) and from real data (green circles)*

The emitter coefficient for this engineered event was determined to match the measured flow rate for the engineered event of approximately 0.6 l/s. From Figure 6-11 the difference between the magnitude of the changes in pressure was readily apparent. On average, the hydraulic model was much less sensitive to the event than sensors in the real DMA were. One potential cause for this behaviour was that there was some change which occurred in the DMA or the WDS upstream of it which caused widespread changes in pressure in the DMA. All sensors, excluding sensor 6 which is located downstream of a PRV, were significantly affected by engineered event number 8. Another potential cause is that the hydraulic model does not exactly match the real DMA. The changes in pressure are dependent upon the changes in flow in all pipes throughout the DMA and differences in certain model parameters, for example valve position, pipe roughness or customer demand, and their equivalents in the real DMA would cause the additional flow to take a different route through the DMA. There were two other engineered events, numbered 5 and 16, at the same location which caused similar changes in pressure for all sensors throughout the DMA as engineered event 8. The similarity between the three different engineered events at this location, which were conducted over the course of several weeks, indicated that the difference between the hydraulic model and the engineered events was caused by differences between the characteristics of the hydraulic model and the real DMA.

To allow a better comparison of the distribution of changes in pressure for each of the sensor locations, the real and modelled data plotted in Figure 6-11 were rescaled so that each fall between the values of 0 and 1. The rescaled real and modelled data are plotted in Figure 6-12.

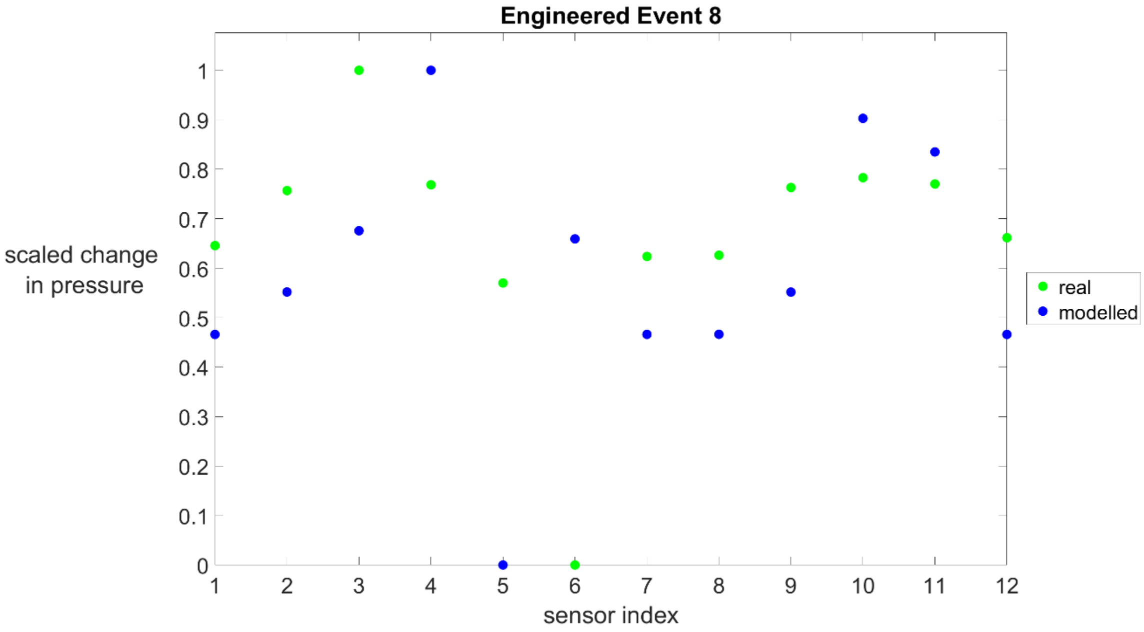


Figure 6-12 *Rescaled changes in pressure for engineered event 8 determined using the hydraulic model (blue circles) and from real data (green circles)*

For each set of changes in pressure plotted in Figure 6-12 this meant that the smallest value equalled 0 and the largest equalled 1. The remaining values were scaled proportionally to their values between the largest and smallest values. The combination of the rescaled data and the raw data allowed several trends to be observed. Two groups of sensors located in the same two looped sections of the DMA exhibited consistent behaviour. The first group of sensors, located in a looped section close to the inlet of the DMA were 1, 7, 8 and 12. The raw and scaled changes in pressure were all similar for these sensors. The second group of sensors were 2 and 9 which were located in a looped section close to the event locations but fed from a different point on the main pipe coming into the DMA.

Sensor 3 registered the highest change in pressure although this sensor was far from the event location. By comparing the change in pressure at sensor 3 with the changes at sensors 10 and 11, which were nearer to event location, it is likely that some change, which is most likely to be an increase in demand relative to the previous days in the vicinity of sensor 3 was responsible for this. Sensor 6, which was close to sensor 3 although separated by a PRV, registered almost no change in pressure which was significantly lower than the modelled change in pressure. A very similar picture was seen for the other two engineered events which were conducted at this location. The lower pressure upstream of the PRV (near to sensor 3), which governs the position of the PRV, would cause the valve position to change which affects the downstream pressure where sensor 6 is located.

The normalised, rescaled changes in pressure, determined by using the normalisation technique described in section 4.4.1.3, were plotted for both the real and modelled data in Figure 6-13. The rescaling procedure was applied to the data plotted in Figure 6-13. The normalisation allows the average sensitivity of the sensor locations to be accounted for so that the changes in pressure are expressed as a multiple of the average sensitivity. In relation to engineered event 8, the normalisation ensured that sensor 4, which was the closest sensor to the engineered event location, was jointly ranked as having the highest normalised change in pressure. This meant that, along with the area around sensor 3, the engineered event location was included in the search area which covered just 14% of the DMA which is plotted in Figure 6-14. The search area for engineered event 8 shown in Figure 6-14 was determined using 10 optimal sensors.

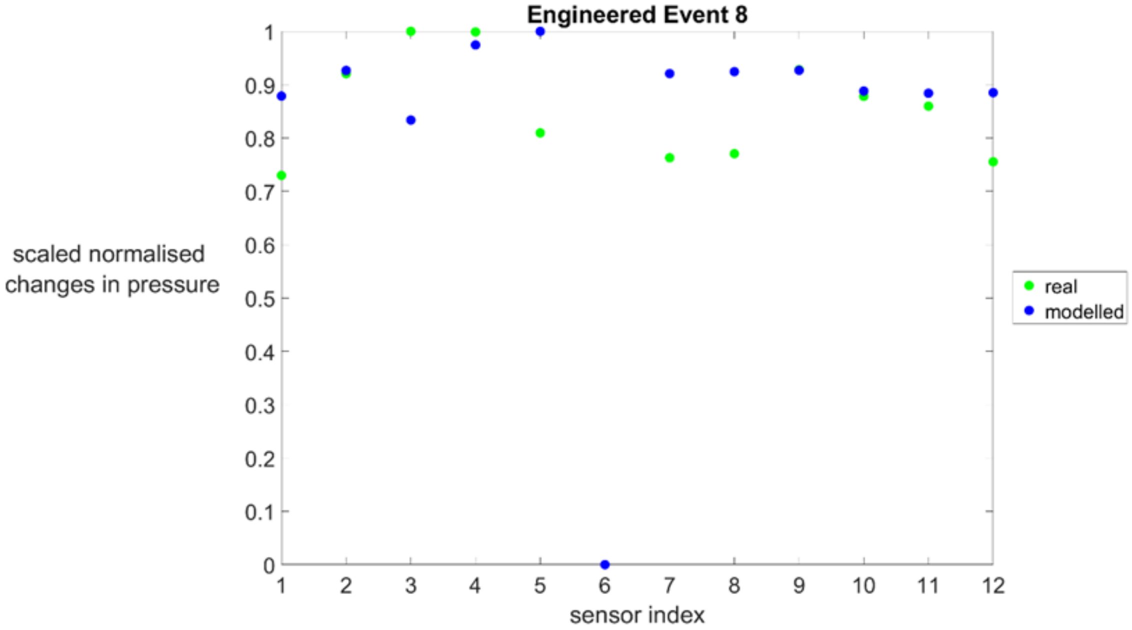


Figure 6-13 *Scaled normalised changes in pressure for engineered event 8 determined using the hydraulic model (blue circles) and from real data (green circles)*

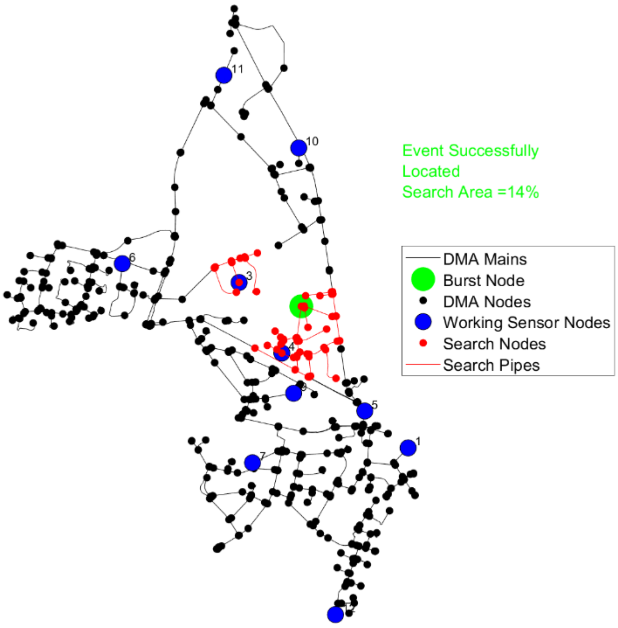


Figure 6-14 *Search area produced for engineered event 8 using the optimal sensor configuration with 10 sensors and the engineered event data*

6.3.2.6. Comparison of agreement between real and modelled localisation results

In addition to comparing the changes in pressure derived from the hydraulic model and the engineered event data collected from the real DMA, a comparison between the localisation performance was conducted for engineered event 8. In Figure 6-15, the size of the search area produced using the modelled data has been plotted on the x-axis and the size of the search area determined using the engineered event data was plotted on the y-axis. A dashed black line, representing the line of perfect agreement between the modelled and real performance is also shown. The proximity of each of the optimal and baseline sensor configurations to the black line denotes the level of agreement between the two. In the ideal case, the size and location of the search area would be identical for both the model and the real engineered event when the same number of sensors were considered. Due to differences between the real DMA and the hydraulic model some difference between the size of the search areas for the model and the real data were expected.

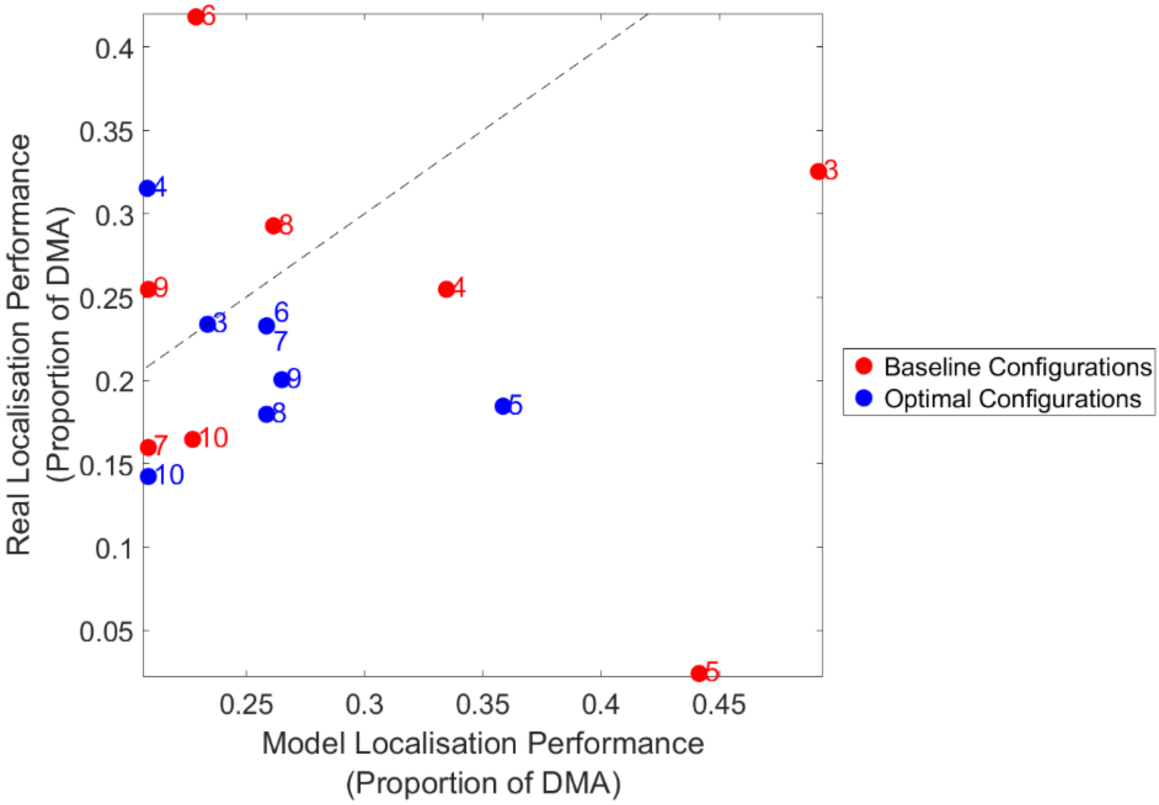


Figure 6-15 *Agreement between the modelled and real search area sizes for engineered event 8 using the optimal and baseline sensor configurations*

Of the optimal and baseline sensor configurations only the optimal sensor configuration with 3 sensors showed perfect agreement between the search areas determined using the hydraulic model and the engineered event. The agreement for this sensor configuration was perfect even though there was a difference between the relative changes in pressure determined using the hydraulic model and for the engineered event. This demonstrates how the leak/burst localisation technique can be used even when a well calibrated hydraulic model is not available for use by the sensor placement technique.

The level of agreement, denoted by the proximity of each configuration to the line of perfect agreement, for the optimal sensor configurations was higher than for the baseline sensor configurations. All optimal sensor configurations, aside from the configuration with 4 sensors, sit below the line of perfect agreement which means that the leak/burst localisation performance was better using the real data than for the modelled data. This is more desirable than the opposite case, where the performance using the modelled data is better than the real data, because the ultimate measure of performance is how well each sensor configuration performs on real data. The performance using the real data is more closely related to the amount of benefit which can be achieved from using the framework of methods with real WDSs.

6.3.2.7. Performance comparison considering only the engineered event locations using both modelled and real data

One key factor which influenced the leak/burst localisation results in Table 6-3 and Table 6-4, as well as the results presented in this section, for the optimal and baseline sensor configurations was that the optimal and baseline sensor configurations were determined using a range of leak/burst event scenarios. It was not feasible to perform the same number of engineered events in a DMA to validate any leak/burst localisation technique or sensor placement technique. Necessarily, a small subset will always need to be used and the achievable performance on the subset of leak/burst event scenarios was very unlikely to be the same as for the set of all possible leak/burst event scenarios. It was, therefore, important to determine what the expected level of leak/burst localisation performance was for the engineered events, in the same way as for the sensor placement technique, so that they could be compared.

To do this, the search areas determined by the leak/burst localisation technique using the modelled data for each of the engineered events, were corrected to account for the two penalties applied within the sensor placement objective function. The same correction was also applied to the search areas determined using the real engineered event data. The Pareto front obtained by running the constrained sensor placement technique (constrained to the deployed sensor locations) was plotted against the corrected average search areas for the constrained optimal and baseline sensor configurations in Figure 6-16. The Pareto fronts, determined by the sensor placement technique considering all of the leak/burst event scenarios determined by the leak/burst event grouping procedure, were plotted in black.

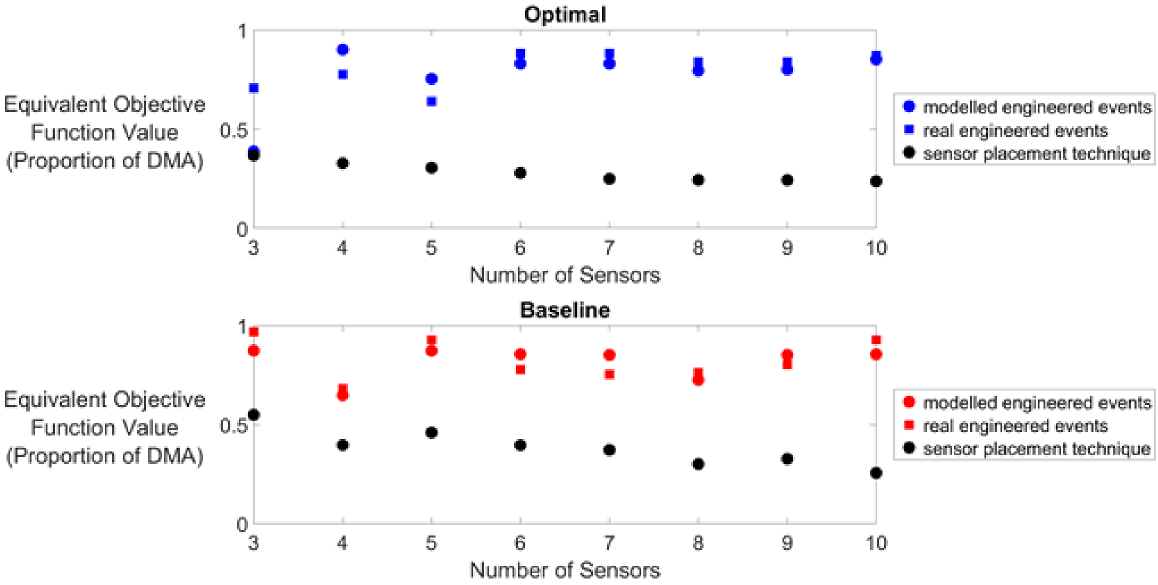


Figure 6-16 *Comparison of the equivalent objective function values for the baseline sensor configurations*

The first clear trend is that the leak/burst localisation performance achieved for the constrained optimal sensor configurations agreed very well when the modelled engineered events were compared to the real engineered events. Aside from the configuration with 3 sensors, the remaining constrained optimal configurations produced average search areas which were within 10% of each other. As the number of sensors was increased the agreement between the results also improved. This is an important result as it demonstrated that comparable performance was achieved even when a water industry standard hydraulic model was used to model the engineered events.

By comparing the sensor placement results, plotted in black, with the modelled engineered events it was clear that the selected engineered event locations and sizes were much more difficult to localise. The only possible reason for this, aside from the different leak/burst event sizes and locations, was that the results plotted in blue were determined using the constrained optimal configuration (selecting from 12 locations) whereas the results plotted in black were determined using the sensor placement technique selecting from 145 hydrants. Further investigation to quantify the effect of the leak/burst event sizes is presented in Figure 6-17. The level of constraint, and its effect on the sensor placement and leak/burst localisation techniques is presented in the next section.

The final important result is that the integration of the sensor placement technique and the leak/burst localisation technique could not fully validated due to the limitations with the available pressure data. In particular, limiting the sensor placement technique to consider only the 12 deployed sensor locations meant that there was a high degree of similarity between the constrained optimal and baseline sensor configurations. The fact that the performance achieved using the modelled engineered events for the baseline sensor configurations agreed more closely to the sensor placement technique than for the constrained optimal configurations meant that the baseline configurations were more favourable for localising the set of engineered events.

6.3.2.8. Effect of Constraint on Sensor Placement and Leak/Burst Localisation Performance

The constraint included in the sensor placement technique, which limited the available sensor locations to only hydrants, has a limiting effect on the sensor configurations which can be made and on the sensor placement performance which can be achieved as a result. This is also true for the constraint due to the deployed sensor locations which were used to validate the framework of methods. Therefore, a comparison between the Pareto fronts for several different versions of the sensor placement technique, with varying levels of constraint on the sensor locations, was conducted to determine how the level of constraint affects the achievable performance of the sensor placement technique. Three different levels of constraint were applied to the sensor placement technique which are listed below.

- No constraint – all hydraulic model junctions were valid locations for placing sensors
- Constrained to hydrants – only junctions corresponding to the location of hydrants were valid locations for placing sensors
- Constrained to deployed sensor locations – only the locations of deployed sensors were valid locations for placing sensors, as per the results in Table 6-3 and Table 6-4, described in section 5.3.2.4.

In Figure 6-17 the sensor placement performance, measured using the average size of the search areas produced, was plotted against the number of sensors for the three different levels of constraint.

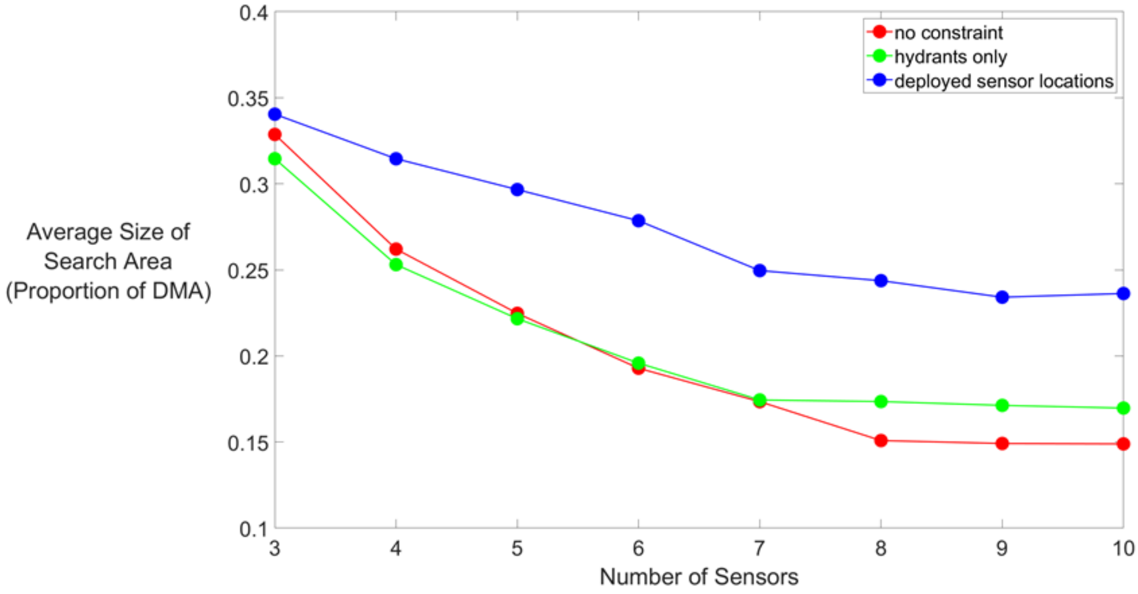


Figure 6-17 *Comparison of the sensor placement performance for three levels of constraint of the sensor placement technique*

The most constrained version, plotted in blue, was limited to just 12 deployed sensor locations and the least constrained version, plotted in red, was able to choose from 731 junctions in the hydraulic model. Between these two, plotted in green, the sensor placement technique was able to choose from 145 hydrants in the DMA. There was a correlation between the level of constraint and the achieved sensor placement performance. The lower the level of constraint, where there were more potential sensor locations to choose from, the better the sensor placement technique performed. The most constrained version of the sensor placement technique which only considered the 12 deployed sensor locations (plotted in blue) performed worse across the range of sensors which were considered. The version with no constraint (plotted in red) performed similarly to the version considering hydrants (plotted in green) for the cases with fewer than 8 sensors but when 8 or more sensors were considered the version with no constraint performed better by between 2-3% of the DMA. This indicated that the constraint of only being able to select from the deployed sensor locations for the purpose of validating the sensor placement technique led to a reduction in the leak/burst localisation performance achieved by the real sensors deployed in the DMA. The constraint of only selecting from the hydrant locations did not lead to a significant reduction in the sensor placement performance which indicates that applying this constraint as part of the sensor placement technique is valid, particularly in light of the ease of installation and the reduced cost of deploying sensors at hydrants.

Due to the combinatorial nature of the sensor placement problem, the effect of the level of constraint on the sensor placement performance was related to the number of sensors being considered. Generally, the fewer sensors considered the less effect that the level of constraint had on the sensor placement performance. Considering the highest level of constraint, deployed sensor locations which was plotted in blue in Figure 6-17, the number of possible configurations with 10 sensors was 66 whereas for configurations with 3 sensors there were 220 possibilities. When the constraint was relaxed (considering hydrants only) the number of possible configurations with 10 sensors was approximately 8×10^{14} and for configurations with 3 sensors there were approximately 500,000 possibilities. The level of constraint affected whether there were more possible combinations of sensors using 3 sensors or 10 sensors. A reduction in the number of potential sensor locations does not lead to the same amount of reduction in the number of possible sensor configurations. The performance of the sensor placement technique is extremely likely, although this has not been proven, to be non-linear against the number of possible sensor configurations to choose from.

To quantify the effect of the different levels of constraint on the validation results in Table 6-3, the leak/burst localisation technique was performed for each modelled engineered event using the optimal sensor configurations and combinations of parameters determined using each level of constraint plotted in Figure 6-17. The average size of the search areas produced, including the penalties for incorrect localisation and multiple search areas, for all 16 engineered events was determined for each number of sensors. The results for the three level of constraint are plotted in Figure 6-18.

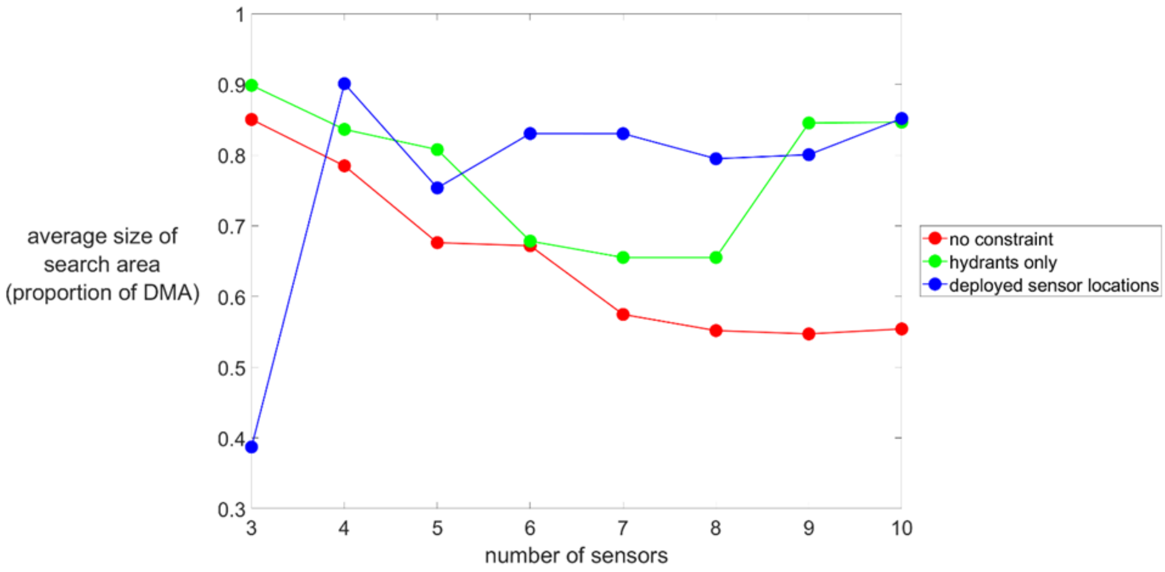


Figure 6-18 *Comparison of the leak/burst localisation performance for the 16 modelled engineered events for three different levels of constraint of the sensor placement technique*

The first important observation is that the least constrained version of the sensor placement technique, where all hydraulic model junctions were considered as valid sensor locations, performed best overall. Aside from the case with three sensors, smaller average search areas were produced for all sensor configurations than for either of the two other constraint levels. As the constraint was increased, by considering all hydrants as valid sensor locations, a marked decrease in the leak/burst localisation performance was seen. The most constrained version of the sensor placement technique, which considers only the 12 deployed sensor locations, was, on average the worst performing. By comparing the results in Figure 6-17 with those in Figure 6-18 the effect of only considering a small number of leak/burst events was clear. The localisation results in Figure 6-18 did not produce smooth lines which was because the smaller number of event locations considered was much more sensitive to individual sensor locations. The constrained optimal and baseline sensor configurations were not determined with respect to only the engineered events therefore the results in Figure 6-18 are less smooth than those presented in Figure 6-17.

By comparing the green and blue lines in Figure 6-18 the effect of constraining the sensor placement technique to only selecting from the deployed sensor locations was demonstrated. On average, the effect was to reduce the leak/burst localisation performance by approximately 1% which was not significant however this was mainly because the blue sensor configuration with 3 sensors was offsetting the reduction in leak/burst localisation performance. For the sensor configurations with more than 3 sensors the average reduction in localisation performance was approximately 6% of the size of the DMA.

6.3.2.9. Factors affecting leak/burst localisation success and accuracy

From the leak/burst localisation results presented in Table 6-3 (optimal sensor configurations) and Table 6-4 (baseline sensor configurations) several important factors were identified which influenced the success and accuracy of the leak/burst localisation technique. Further investigation of these factors is presented in this section to examine the extent to which each contributes to the success of the leak/burst localisation technique.

Distance to the nearest sensor for each event

In Figure 6-19 the distances to the nearest pressure sensor were plotted against the size of the search area produced for each of the correctly localised engineered events. The issue of proximity to pressure sensors is pertinent because this can influence whether the largest changes in pressure due to a leak/burst event are likely to be captured by a configuration of sensors. Using the data plotted in Figure 6-19, a weak correlation (ρ = 0.249) was found indicating that there was minimal statistical significance between the distance to the nearest sensor and the size of the search area produced.

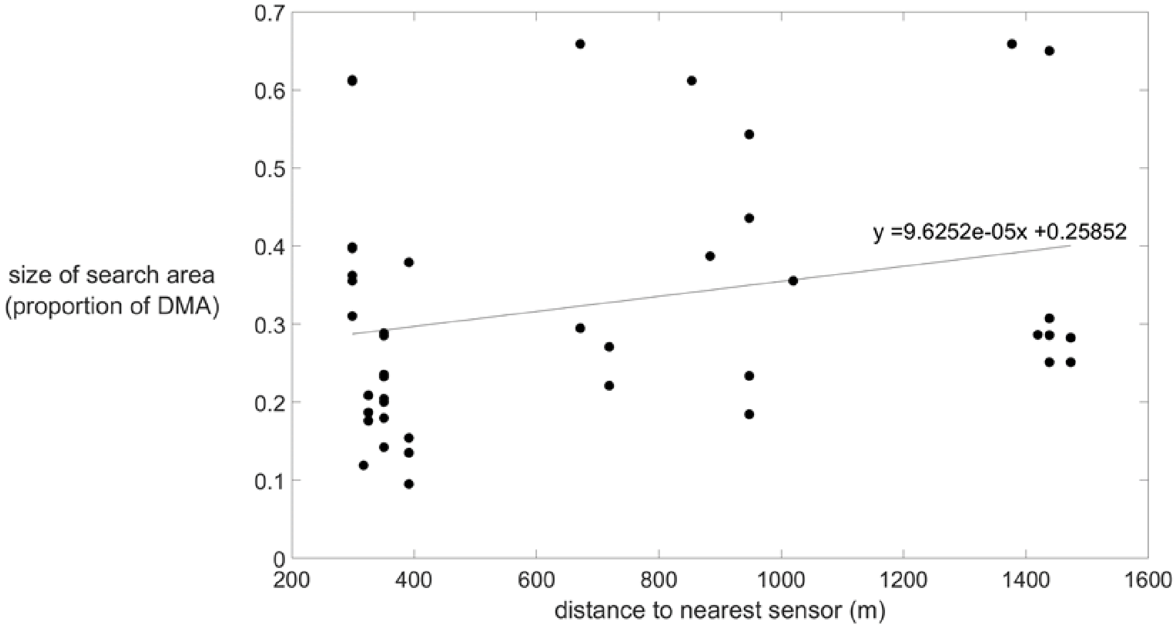


Figure 6-19 *Comparison of the distance to the nearest sensor and the size of the search area produced for all successfully localised events considering all optimal sensor configurations*

There is an important reason why this was the case. Firstly, the distance to a single sensor is more important in terms of detection rather than localisation. This is because detection only requires that a single sensor captures a large enough change in pressure which can be distinguished from the normal variation which occurs even when no leak/burst events are present. To localise a leak/burst event, the combination of several changes in pressure are normally used (as per the leak/burst localisation technique developed and presented in this thesis) meaning that the significance of the proximity to a single sensor is less important. In addition to this, the average distance to the nearest sensor was 649 metres for the engineered events which were correctly localised compared to 604 metres for the events which were not correctly localised. This indicated that the distance to the nearest sensor alone was not responsible for events being correctly localised.

Small size of the engineered events

The sensor placement technique automatically determines the size of the leak/burst events and groups them prior to determining the optimal sensor configurations and leak/burst localisation parameters. This means that the optimal sensor configurations and leak/burst localisation parameters are optimal with respect to the range of leak/burst event sizes which were determined. The size of the real-world engineered events was limited to a fixed flow rate of approximately 0.6l/s in order to prevent problems with customer supplies in DMA 304-03. This was approximately 50% of the smallest leak/burst events which were considered by the sensor placement technique for this DMA. The effect of this is demonstrated in Figure 6-20. The average leak/burst localisation performance for all 16 modelled events was plotted for various leak/burst event sizes using the optimal sensor configurations and parameter combination from the 12 deployed sensor locations. The smallest size plotted was the actual event size and then each of the leak/burst event sizes determined by the sensor placement technique (with emitter coefficients ranging from 0.1 to 0.7) have also been plotted.

Figure 6-20 demonstrates that the small size of the engineered events did negatively impact the leak/burst localisation performance when compared to the range of event sizes generated by the sensor placement technique. The general trend was that as the leak/burst event size increased the average size of the search area decreased across the 16 modelled engineered events. The emitter coefficient of 0.5 produced the smallest average search areas and further increases in event size above this did not lead to further reductions in the average search area size. The actual size of the engineered events led to the significantly worse performance than for any of the event sizes determined by the sensor placement technique. Comparing the actual event sizes (black line with crosses) with the worst performing of the determined leak/burst event sizes (C = 0.1 - red line with crosses) the average reduction in performance was on average approximately 6% of the DMA. For the best performing of the determined leak/burst event size (C = 0.7 – blue line with circles) the reduction in leak/burst localisation performance was 47% of the DMA on average. This demonstrated that the small size of the leak/burst event was more impactful than using the constrained version of the sensor placement technique.

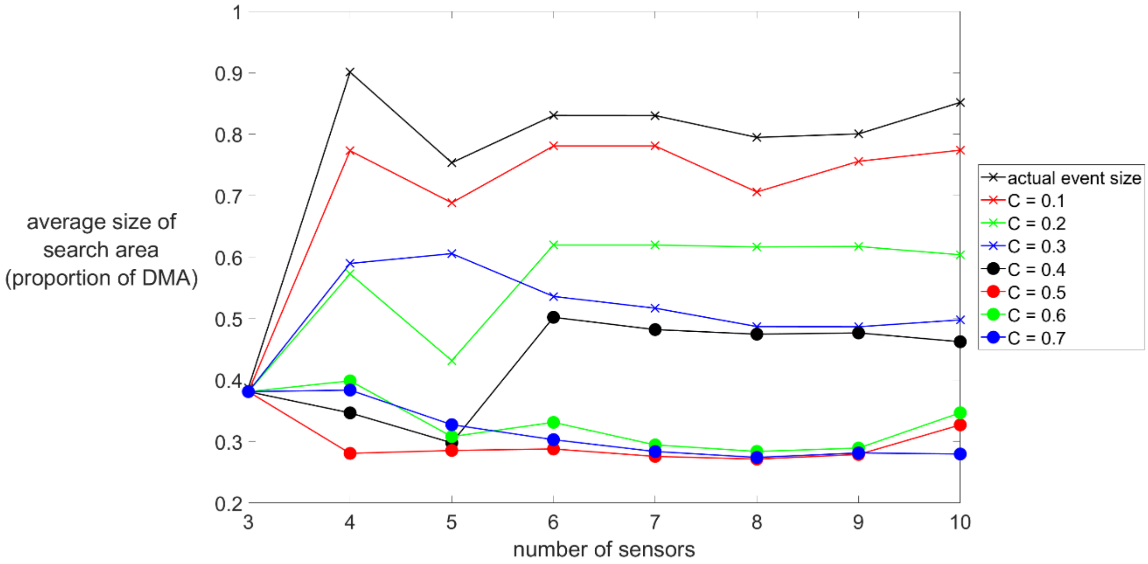


Figure 6-20 *Comparison of the leak/burst localisation performance for 16 modelled events considering different sizes of event using the constrained optimal sensor configurations*

<u>Number of engineered events</u>

In total, 16 engineered events were conducted in DMA 304-03 over a period of one month between 28/01/2020 and 28/02/2020. The leak/burst localisation methodology ideally uses the most recent days prior to a leak/burst event to ensure that the operating conditions in the WDS most closely match those on the day of the leak/burst event, aside from the presence of the leak/burst event itself. Previous event days were removed from the previous day data set to ensure that the changes in pressure due to other events are not included in the leak/burst localisation. Therefore, with so many events in a short space of time, many of the most recent days were removed from the previous day data set and days from further in the past were included to allow the leak/burst localisation to be performed. By examining the previous day data sets with the dates of the engineered events a maximum difference of 39 days was found for engineered event 16. When more than two weeks of data was used during the development of the leak/burst localisation technique a significant impact upon the leak/burst localisation accuracy was found. Therefore, by performing the engineered events in this way, and having to include older data in the previous day data set, it was expected that the localisation accuracy would be negatively impacted.

The results for the second validation DMA, DMA 304-04, are presented in the next section.

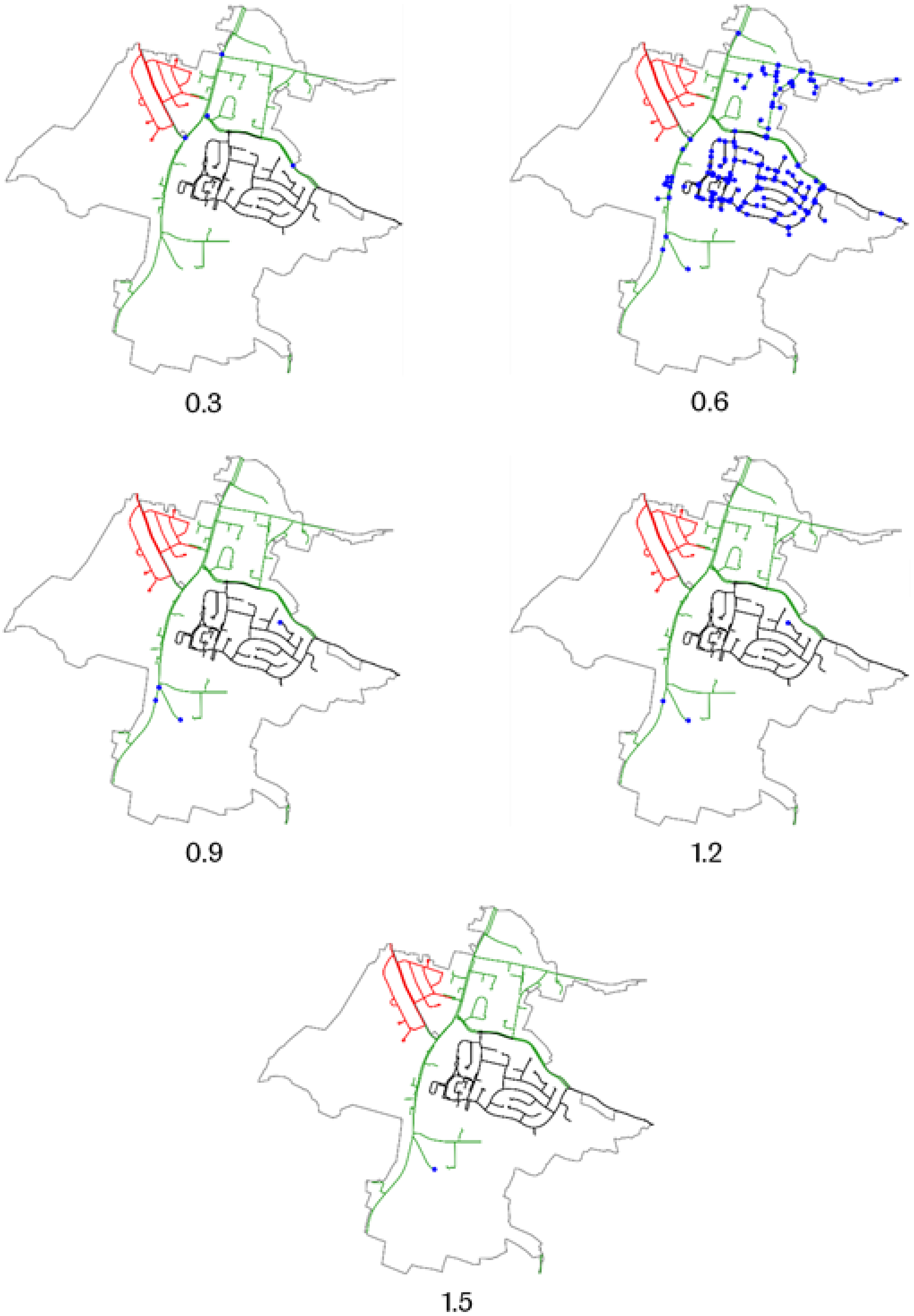
6.3.3. DMA 304-04

6.3.3.1. Leak/Burst Event Sizes and Leak/Burst Event Grouping

Table 6-5 shows the number of groups of leak/burst events for each of the leak/burst event sizes considered for DMA 304-04. An event size increment of 0.3 and a maximum allowable flow increase of 0.1 were used to determine the range of leak/burst event sizes. The largest leak/burst event size for which valid groups were produced was an emitter coefficient of 1.5. The same trend in the number of groups of leak/burst events for each leak/burst event size was seen for DMA 304-04 as for DMA 304-03. There were only a small number of leak/burst events in all groups aside from the group for an emitter coefficient of 0.6. A total of 272 leak/burst event scenarios were considered by the sensor placement technique. Comparing this with the 706 leak/burst event locations and the 5 leak/burst event sizes the reduction in the number of considered leak/burst event scenarios was 92.3%. In Figure 6-21 the leak/burst event locations for each size of event used by the sensor placement technique are shown.

Table 6-5 *Leak/burst event sizes and valid leak/burst event grouping results for DMA 304-04*

Leak/Burst Event Size	**Number of valid leak/burst event groups**
0.3	6
0.6	257
0.9	5
1.2	3
1.5	1
Total	**272**



0.3

0.6

0.9

1.2

1.5

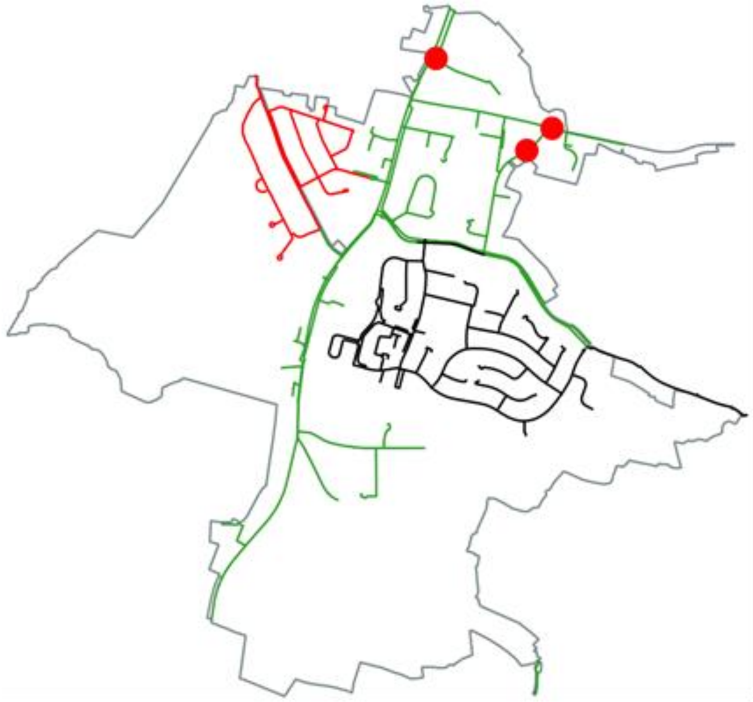
Figure 6-21 *Grouped leak/burst event locations for each of the emitter coefficients determined by the automatic leak/burst event size determination procedure for DMA 304-04*

6.3.3.2. Optimal and Baseline Sensor Configurations and Parameters

The optimal and baseline sensor configurations and parameters for DMA 304-04 are shown in Table 6-6 and Figures 6-22 and 6-23. In this case, the sensor placement technique only determined the optimal configuration with 3 sensors. This was because no configuration with more than 3 sensors performed better than the optimal configuration with 3 sensors. The GALAXY MOEA will tend to remove members of the population which perform worse in terms of the number of sensors and the leak/burst localisation performance. There is no guarantee that the Pareto front determined by the sensor placement technique will find optimal solutions for the entire range of numbers of sensors. The range of the number of sensors in the determined optimal sensor configurations is DMA specific and in some cases, such as for DMA 304-04, the maximum number of sensors to be deployed will not equal the maximum allowable number of sensors, which is set to 10 for the sensor placement technique. All baseline sensor configurations and parameter combinations with between 3 and 10 sensors (inclusive) are shown in Table 6-6 and Figure 6-23.

Table 6-6 *Leak/burst localisation parameters determined for the baseline and optimal sensor configurations for DMA 304-04*

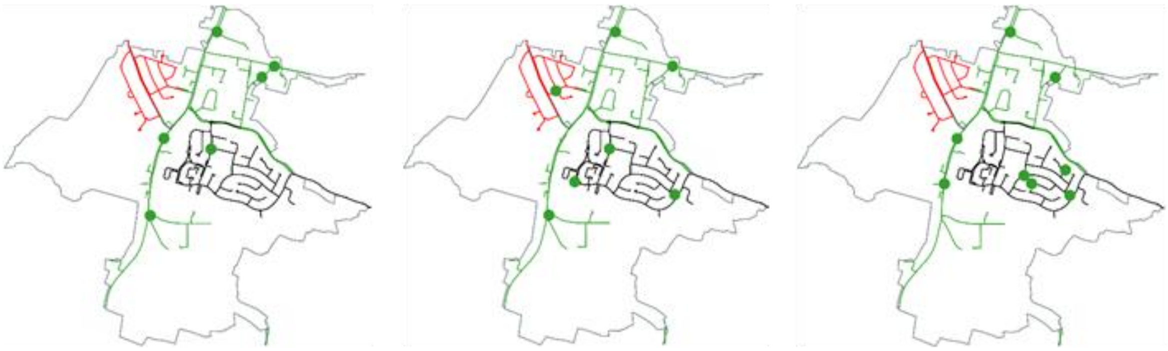
Number of Sensors	**Baseline**		**Optimal**	
	Search area threshold	**Interpolation exponent**	**Search area threshold**	**Interpolation exponent**
3	0.5	1	0.95	30
4	0.5	6	-	-
5	0.85	25	-	-
6	0.6	28	-	-
7	0.85	12	-	-
8	0.85	8	-	-
9	0.95	25	-	-
10	0.9	25	-	-



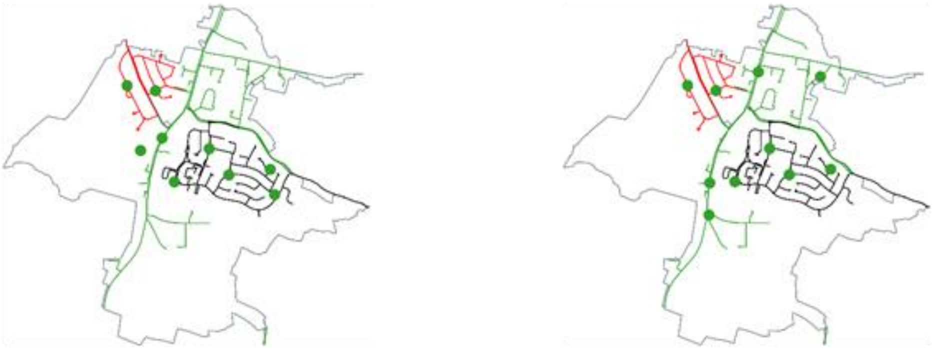
3 Sensors

Figure 6-22 *Optimal sensor configuration (red circles) with 3 sensors for DMA 304-04*

3 Sensors 4 Sensors 5 Sensors



6 Sensors 7 Sensors 8 Sensors



9 Sensors 10 Sensors

Figure 6-23 *Baseline sensor configuration (green circles) for DMA 304-04*

6.3.3.3. Comparison of Optimal and Baseline Sensor Placement Performance

The value of the objective function determined by the sensor placement technique for both the optimal and baseline sensor configurations was plotted in Figure 6-24. For DMA 304-04 only the sensor configuration with 3 sensors was identified. This was because no sensor configurations comprised of more sensors was able to perform better than the configuration with 3 sensors. Therefore, the sensor placement technique discarded the configurations with higher numbers of sensors. The baseline configurations for all numbers of sensors are shown although only the baseline configuration with 3 sensors will be considered from this point forward due to the lack of optimal configurations with higher numbers of sensors against which to compare their performance. The performance for the optimal sensor configuration was approximately twice as good as the baseline sensor configuration as determined by the sensor placement technique.

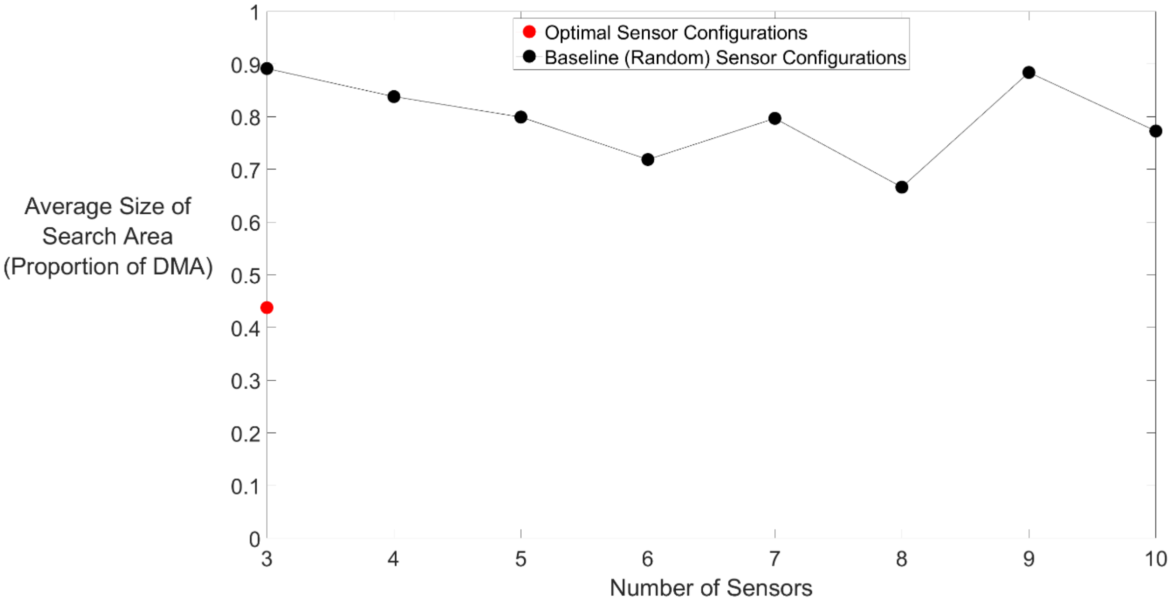


Figure 6-24 *Comparison of the performance for the optimal and baseline sensor configurations for DMA 304-04*

6.3.3.4. Comparison of the leak/burst localisation performance using the engineered event data

The optimal and baseline configurations were used to localise the four engineered events which were performed in DMA 304-04. The same procedure described in section 6.3.2.4 for DMA 304-03 was used to evaluate both sensor configurations for DMA 304-04. The localisation results are given in Table 6-7 below for each engineered event.

Table 6-7 *Results for the optimal and baseline sensor configurations for the 4 engineered events in DMA 304-04*

Engineered Event Number	**Optimal Sensors**	**Baseline Sensors**
	3	**3**
1	49.2	27.3
2	30.8	26.9
3	**50.5**	26.9
4	**38.0**	**35.3**
Average Search Area	42.1	29.1
Correct Events	2	1
Events with multiple search areas	0	0
Equivalent objective function value	72.1	83.8

The optimal sensor configuration successfully localised two of the engineered events compared to only one event for the baseline configuration. The average search area produced by the optimal sensor configuration was larger than for the baseline sensor configurations although the equivalent value of the objective function (including penalties) was lower for the optimal configuration. Neither sensor configuration produced any events for which multiple search areas were determined. This is typical for cases where fewer sensors were used as the resulting surface produced by the leak/burst localisation is usually smoother.

6.3.3.5. Comparison of performance considering only the engineered event locations using both modelled and real data

In Figure 6-25, a comparison of the leak/burst localisation performance achieved using both the constrained optimal and baseline sensor configurations was plotted. For each sensor configuration the achieved sensor placement performance was plotted as a black circle. The equivalent performance, achieved considering only the engineered event locations and sizes in the hydraulic model, were plotted as coloured circles. The equivalent performance considering the actual engineered events was plotted as a coloured squared.

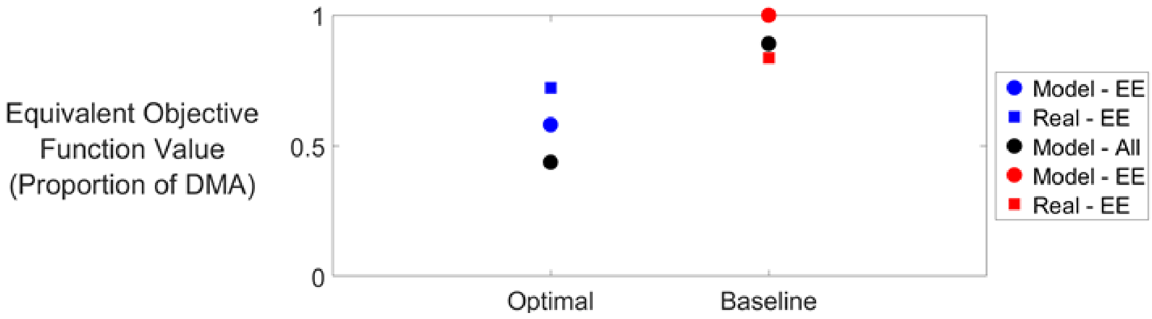


Figure 6-25 *Comparison of the leak/burst localisation performance achieved using the optimal and baseline sensor configuration*

Although only a partial trend could be seen due to only considering one baseline and one constrained optimal sensor configuration the behaviour of DMA 304-04 was different to DMA 304-03. Firstly, the difference between the equivalent performance on the modelled engineered events (coloured circles) and using the sensor placement technique (black circles) was similar for both sensor configurations. The baseline sensor configuration performed approximately 11% worse considering the modelled engineered events and the constrained optimal sensor configuration performed 15% worse. This showed, on average, the engineered events scenarios were more difficult to localise than the set of leak/burst event scenarios considered by the sensor placement technique. In this DMA only four engineered events were performed. Ideally, selecting engineered event locations at random, as per the Monte Carlo approach used by Hagos et al. (2016) would select a sample of leak/burst event scenarios which are representative of all possible leak/burst event scenarios. However, the number of events selected must be sufficiently large for the behaviour of the sample of engineered events to be representative of the overall population of all leak/burst event scenarios. There is a practical limitation of the number of engineered events which can be conducted in a DMA, to minimise the amount of time and resources which are required to conduct them and to minimise the risk to customer supplies which can be caused by performing engineered events.

The equivalent localisation performance determined using the real engineered events differed significantly between the baseline and optimal sensor configurations. The baseline sensor configuration performed better using the real engineered event data than the modelled event data by approximately 16% of the DMA on average. The optimal sensor configuration, however, performed worse by approximately 14% of the DMA. Given the uncertainty which is inherent in a real WDS and which does not exist in a hydraulic model it is logical that perfect agreement between the localisation performance was not achieved. Unpredictable events in the DMA, for example an unusually large demand in an area of the DMA which does not contain the engineered event could easily skew the results, particularly in the case for this DMA where the size of the engineered events was equal to approximately 4.6% of the average daily rate of flow into the DMA calculated over a normal week where no leak/burst or abnormal flow events were present. However, given that the size of the engineered events was larger than would be expected from a single household or even several households it would be extremely unlikely for an event (or multiple events in the same area) to occur which would mask or skew the effect of the engineered events.

As part of the engineered event trials, engineered events were also conducted in DMA 304-06 which sits downstream of DMA 304-04 on the same nights as the engineered events in DMA 304-04. DMA 304-06 was not included in the validation DMAs due to lower number of pressure sensors deployed in this DMA and problems with the data from those sensors. The impact of the engineered events in DMA 304-06 on the pressure sensors in DMA 304-04 was noticeable. DMA 304-04 contained 3 DPAs, two of which were protected from excess pressure by PRVs. The pressure sensors located in the largest DPA in DMA 304-04 showed, on average, larger changes in pressure across the four engineered events. This was true even when the engineered events were located in one of the smaller DPAs and would, therefore, not be expected to have a significant effect on the sensors in the largest DPA. This demonstrated that the engineered events in DMA 304-06 were affecting the sensors in DMA 304-04 to such an extent as to cause difficulties in localising the engineered events in DMA 304-04.

6.4. Chapter Summary

In this chapter the framework of methods, which was described in Chapter 3, has been validated by applying it to a real DMA in United Utilities WDS. The sensor placement technique was run using the hydraulic model of the DMA to determine the optimal (constrained) sensor configurations from 12 available sensors which were deployed in the DMA. 16 engineered events were performed in 8 different locations throughout the DMA were used to test how well the optimal sensor configurations performed in comparison to the baseline configurations. In total the optimal sensor configurations successfully localised more of the engineered events with smaller search areas, on average, than the baseline sensor configurations.

The effect of heavily constraining the sensor placement technique, due to uncontrollable circumstances limiting the deployment of sensors throughout the DMA, was demonstrated using modelled engineered events for DMA 304-03. The limitation on the sensor locations most significantly impacted the optimal sensor configurations with higher numbers of sensors, when all modelled leak/burst events were considered. Additionally, the performance using the engineered events was better for optimal sensor configurations with 5 or fewer sensors. Due to the constrained sensor placement, the expected difference between the leak/burst localisation for the constrained optimal and baseline sensor configurations was limited. Reducing the level of constraint, by considering all hydrants as valid sensor locations as per the proposed sensor placement technique, as demonstrated in Figure 6-17.

Furthermore, a comparison of the localisation accuracy achieved using the baseline and constrained optimal sensor configurations, demonstrated that the effect of the chosen engineered event locations significantly impacted upon the leak/burst localisation performance achieved using the optimal sensor configurations. The baseline sensor configurations were expected to outperform the optimal sensor configurations in most cases due to the number and location of the selected engineered event locations. This was caused because the leak/burst event scenarios used to determine the optimal sensor configurations differed significantly in the total number of events and the sizes of the events. However, the optimal sensor configurations tended to outperform the expected modelled leak/burst localisation performance when the engineered events were considered.

The validation case study was performed using small engineered events which were less than 5% of the average flow rate for both of the validation DMAs. Using engineered events of this size had the effect of producing small differential changes in pressure measured at the sensor locations throughout each validation DMA. For many of the engineered events, the range of changes in pressure, considering all sensor locations, were less than the accuracy of the instrument. This meant that the effect of those events was not adequately captured by the deployed sensor locations leading to one of two outcomes. The first outcome was that the event was incorrectly localised, caused by a sensor far away from the event registering a larger change in pressure than the sensors closer to the event. The second outcome was that a larger search area was produced because sensors far away from the event were also included in the search area along with those which were near to the event. The small size of the engineered events was demonstrated to have a significant negative impact of the leak/burst localisation performance achieved using the constrained optimal sensor configurations.

For a single engineered event in DMA 304-03 the agreement between the search area sizes produced for the optimal and baseline sensor configurations demonstrated that the optimal sensor configurations gave better agreement. For the optimal sensor configuration with 3 sensors, which was correctly localised using the real and modelled data, exactly the same search area was produced even though the differential changes in pressure for that event differed.

Chapter 7 – Discussion

7.1. Introduction

A framework of methods has been developed to localise new leak/burst events in WDS using a data-driven leak/burst localisation technique integrated with a sensor placement algorithm. The leak/burst localisation technique is used by the sensor placement technique ensuring that the determined sensor configurations are optimal with respect to the leak/burst localisation technique. Additionally, multiple leak/burst event sizes and locations are considered ensuring that the range of leak/burst event scenarios of interest, namely those events which are typically unreported and cause changes in pressure throughout a DMA, are used to determine the optimal sensor configurations.

7.2. Key Items

7.2.1. Integration of leak/burst localisation and sensor placement

The determined optimal sensor configurations directly consider the leak/burst localisation technique being used. Leak/burst localisation techniques have different performance characteristics dependent upon the analytical methods they are comprised of and no sensor configuration can be considered as optimal for a DMA independent of the leak/burst localisation technique. Sensor placement approaches do exist which are only dependent on spatial considerations in a DMA, such as that proposed by Christodoulou et al. (2013). Aside from discounting the hydraulics of a DMA, which is a critical omission, no leak/burst localisation was considered either.

A central part of the integrated framework is the use of the localisation accuracy, measured using the length of pipes (as opposed to the number of junctions) in the search area as a proportion of the DMA, which is minimised by the sensor placement technique directly. In addition to this, an identical analytical procedure is used to determine the search area for both the sensor placement technique and the leak/burst localisation technique. The only difference is that real data is used by the leak/burst localisation technique whereas hydraulic model simulated data is used by the sensor placement technique. From the literature on sensor placement techniques, discrepancies between the analytical procedures for sensor placement and leak/burst localisation are common meaning that the sensor configurations are optimal with respect to the sensor placement technique and not necessarily the leak/burst localisation technique. Casillas et al. (2015), for example, used the similarity between the real data and known hydraulic model leak/burst scenarios for leak/burst localisation but optimised the sensor locations to minimise similarity between the modelled leak/burst event scenarios. There is no mechanism to ensure that the number of leak/burst events localised correctly was maximised although this technique only provides a single junction as the suspected leak/burst event location. A similar scenario was seen for the technique developed by Soldevila et al. (2019) which used the maximum difference between two interpolation surfaces (produced using Ordinary Kriging) to identify the leak/burst location. The sensor placement technique, however, minimised the error for the Ordinary Kriging model by comparing it to the modelled leak/burst event scenarios. This does not imply that the correct leak/burst event location will be identified just because the Kriging model is as accurate as it can be.

A range of optimal sensor configurations, in the form of a Pareto front, are identified by the presented sensor placement technique. This allows flexibility and provides multiple solutions which are all optimal with respect to the range of leak/burst event sizes which were modelled and the available sensor locations to choose from. The benefit of this approach is that the knowledge and experience of water company operations staff and decision makers can be used to evaluate all optimal sensor configurations in lieu of a formalised decision support framework. Furthermore, the localisation performance can be considered within the framework of regulatory targets which most water companies in developed countries operate within.

The validation of the framework of methods, presented in Chapter 6, demonstrated that using non-optimal (in that case a randomly generated baseline) sensor configurations led to worse leak/burst localisation performance than using the optimal configurations when modelled and real leak/burst events were considered. The baseline sensor configurations were generated randomly from the available sensor locations meaning that they were not determined with respect to the leak/burst localisation technique, a sensitivity matrix or the leak/burst localisation accuracy. The implication of this is that using another sensor placement technique which does not consider the leak/burst localisation technique will lead to a loss of leak/burst localisation performance.

The integrated framework presented in this thesis can be easily adapted for any data-driven or model-based leak/burst localisation technique such that the localisation accuracy, including the number of leak/burst events correctly localised, is maximised.

7.2.2. Development and validation of framework of methods using a combination of modelled and real leak/burst event data

The sensor placement and leak/burst localisation techniques identified in the literature use a wide range of methods for evaluating their performance. The importance of the type of data used for testing and validation is that real WDS behave differently than even the most accurate/complex hydraulic models. Techniques which are developed using only hydraulic model data, which is common among those approaches identified in the literature, may not perform as well on real WDS and, therefore, in the ideal case, data collected from sensors deployed in real WDSs, should also be used to determine the performance of a particular technique. Another important related issue is that of adoption of new technologies by water companies. Due to the constraint on the available resources for implementing new technologies, water companies will be more likely to adopt techniques which are proven in the most realistic cases because this minimises the risk of investing money in a technology which ultimately does not perform to the required level.

The most common approach in the literature, as was used by Perez et al. (2011) and Casillas et al. (2013, 2015), was to use a hydraulic model to simulate leak/burst events. The performance of those techniques was then evaluated using only the simulated leak/burst events. However, these did not consider any real leak/burst events which implies that the performance which could be achieved using real WDSs is uncertain. The required degree of complexity of the modelled leak/burst events scenarios depends upon the approach being used and varies widely between the techniques in the literature. Elements which can be considered, which were highlighted by Wu and Liu (2017) and Romano (2019), include uncertainty in demand, uncertainty in sensor measurements, failure of sensors and considering multiple leak/burst event sizes. Using a hydraulic model offers the advantage of being able to control the model parameters and to simulate a large number of leak/burst in a short time. An improvement over using a hydraulic model to evaluate performance is to use engineered events, which are conducted in real WDSs by opening fire hydrants. Evaluating performance using engineered events gives greater confidence in the achieved results when related to the achievable performance on real leak/burst events in WDSs. Techniques which have used engineered events to evaluate performance are far less common that those using hydraulic models. Romano et al. (2016) presented by far the largest case study using 132 engineered events in 17 DMAs in a real WDS. Farley et al. (2010) used engineered events in several real DMAs to validate a sensor placement and leak/burst detection/localisation technique. Some techniques were found, for example Sophocleous et al. (2019), which used a real (i.e. occurring naturally and not by opening a fire hydrant) historical leak/burst event to evaluate their leak/burst localisation technique. The general trend is that the performance using engineered or real leak/burst events is typically much worse than when hydraulic models are used. Also, comparing the performance of two different leak/burst localisation techniques is problematic as the data used for each technique validation are usually created specifically for that technique. This means that different sets of data are used for each technique and the technique itself is not isolated as control variable.

The development and validation of the framework of methods used three DMAs within United Utilities WDS. One DMA (DMA 123-09) was used during development, and 2 DMAs (DMA 304-03 and DMA 304-04) were used for validation. Using different DMAs to validate the framework than were used for its development ensured that the performance for DMAs with different characteristics was demonstrated. 31 engineered events were conducted in the three DMAs used during the development and validation of the framework of methods. 11 of these events were used to develop the leak/burst localisation technique and 20 engineered events were used to validate the framework of methods.

7.2.3. Data-driven leak/burst localisation technique which does not require use of a well-calibrated hydraulic model

Hydraulic model-based leak localisation techniques are much more common than data-driven approaches. This is caused by the ease with which models can be manipulated and explored and the historic absence of data collected from WDSs which could be used to aid with the development of data-driven techniques. This is particularly true of pressure data which is typically not monitored throughout DMAs although it is usually collected from one or two points. Model-based leak/burst localisation techniques are highly sensitive to the quality of model being used and will only perform well when a highly calibrated model is available. Within water companies, hydraulic models are readily available although they are of variable quality. Often years can go between any updates being made to hydraulic models and even if they are regularly calibrated by water company staff they are typically not of sufficient accuracy to be used with model-based localisation techniques. These techniques rely on a calibration accuracy of 0.1m whereas water company models typically are calibrated to within 1m (Sophocleous et al., 2019b). The additional cost and resources needed for calibrating a large number of hydraulic models to the required accuracy presents a barrier for these techniques to be deployed by water companies.

The leak/burst localisation technique used by the framework of methods does not compare the model with the real data to localise leak/burst events. Instead, the past behaviour of the real WDS is compared with its behaviour when a leak/burst is present to determine the approximate leak/burst location. This removes the reliance on the accuracy of the hydraulic model to identify the correct area of the DMA which contains the leak/burst event. The hydraulic model does, however, still have an impact indirectly as it affects the sensor locations which are determined by the sensor placement technique.

Rather than using a “hotspotting” technique which defines a very small search area (of only a few nodes or pipes in the DMA) the leak/burst localisation technique determines a sub-region of the DMA meaning it is tolerant to discrepancies between the changes in pressures determined by the model and the measured data collected from the deployed pressure instruments. A similar approach was used by Cuguero-Escofet et al. (2015) but a separate distance-derived cross correlation measure was used in conjunction with a correlation-based localisation technique to widen the search area. Using SC-IDW means that this happens automatically so long as the interpolation exponent and search area threshold are set carefully. Optimising both leak/burst localisation parameters, as part of the sensor placement technique, means that the best performance is achieved automatically and that the parameters are customised with respect to the chosen sensor configuration. This was demonstrated in section 5.4.1 for two simple verification hydraulic models used during the development of the framework of methods.

A characteristic of using data-driven leak/burst localisation techniques such as the one presented in this thesis or the technique proposed by Soldevila et al. (2019) is that only newly occurring leak/burst events, which started after the deployment of the pressure sensors, can be localised. Background leaks which started before pressure sensor deployment cannot be localised using these techniques, however, they will not normally impact upon customer supplies and do not present as great a risk to customer supplies as larger leak/burst events. This is a disadvantage of using a data-driven technique which relies on additional pressure monitoring of DMAs when compared to model-based techniques which solve the inverse problem by comparing a hydraulic model with real data such as Sophocleous et al. (2019a). Other types of leak/burst localisation techniques using for example pressure transients (Colombo et al., 2009) or acoustic/noise loggers (Hunaidi and Wang, 2006) are capable of determining the leak/burst location without consideration of the past behaviour of the WDS notwithstanding the much higher cost of those types of sensors when compared with pressure sensors

By using the novel distance function (calculated using the shortest path along pipe lengths) as opposed to Euclidean distance, as part of the leak/burst localisation technique, more realistic and often smaller search areas were produced than when compared to Romano et al. (2013). The novel distance function is able to distinguish between two pipes which are close together but which are not connected directly to each other which represented a significant improvement over the results which were given in Romano et al. (2013) where pipes which are distant, in terms of the length travelled along pipes, were still included in the area with the highest probability of a leak/burst event. This was demonstrated for an engineered event in Figure 5-2 in section 5.3.1.

7.2.4. Sensitivity to leak/burst events, sensor uncertainty and model uncertainty

The measured changes in pressure, determined by performing engineered events in a total of 3 real DMAs, were large enough even when the leak/burst event flow rate was as low as 3.5% of the peak daily inflow to the DMA. Even when the size of the engineered events was very small, as it was for the validation DMAs, engineered events were successfully localised using as few as 3 sensors to less than one quarter of the DMA. This compares favourably with the technique developed by Soldevila et al., (2019) which localised (to within 200m) a leak/burst event at night of approximately 8% of the peak inflow using 5 sensors. The best performance using the optimal sensor configuration with 5 sensors in DMA 340-03 was for event 8 where a search area of approximately 19% of the DMA was produced. A direct comparison of the results is not possible because Soldevila et al. (2019) did not define a search area but used the distance between the actual leak/burst event location and the suspected location.

The pressure at night is more sensitive to small leak/burst event sizes and even events which are smaller than 5% of the average flow rate in to a DMA caused large enough changes in pressure to be measured, with certainty, by pressure sensors as evidenced by the modelled and real changes in pressure plotted in Figure 6-11 in section 6.3.2.5. Here certainty refers to the fact that the changes in pressure were significantly larger than the accuracy of the pressure instruments. The same events (i.e. hydrant openings) occurring during the day would lead to a smaller flow rate and due to the increase in customer demand the changes in pressure due to these events could not be distinguished from the typically larger pressure variations which occur in the day. This means that smaller leak/burst events will cause changes in pressure which are larger than the uncertainty associated with the pressure instruments. The leak/burst localisation technique used by the framework of methods takes advantage of this because it uses the differential changes in pressure to determine the leak/burst event location. When larger changes in pressure are caused by a leak/burst event the differential change, the difference between the largest and smallest change in pressure, also increases. This will produce more focussed search areas in the cases where the sensor(s) in the vicinity of the leak/burst event register the largest changes in pressure. This was demonstrated by the changes in pressure plotted in red for engineered event 8 in Figure 6-11 in section 6.3.2.5. For this event sensors 4, 10 and 11 which sit in the same section of the DMA as the engineered events determined changes in pressure which were largest, aside from sensor 3. Sensor 3 underwent the largest change in pressure although this was in the vicinity of a PRV which was likely having an impact on this sensor, as it was for sensor 6 which was also close to the PRV.

Another benefit related to this (although it did not drive the decision to use the night for localisation) was that using the night time pressure reduced the risk of causing interruption to customer supplies during the engineered events as the likelihood of customers using water is reduced when compared to the day. Other types of potential disruption to customers/members of the public such as the flooding of footpaths is negated as well.

The lower limit of leak/burst event sizes which can be successfully localised is larger than those which can be detected because most data driven detection techniques consider flow which is more sensitive to new leak/burst events than pressure, which has been used by the leak/burst localisation technique. This can be considered as a limitation of using pressure for leak/burst localisation although using pressure is compelling in light of its much lower cost when compared to flow instruments. Using a single flow instrument, as is normal practise in most WDS is many developed countries, cannot allow localisation as it aggregates the flow for the entire DMA.

Deploying multiple flow instruments is not currently a feasible approach due to the cost that this would involve.

The value of 0.1m for the sensor accuracy is typical for pressure sensors deployed in WDS and is the value used Sophocleous et al. (2019b), Soldevila et al. (2019) and Cuguero-Escofet et al. (2015). The importance of this value is that it governs the smallest change in pressure which can be distinguished from a change which is associated with sensor uncertainty. In practical terms this affects the smallest size of leak/burst events which can be localised by the leak/burst localisation technique. Using more accurate sensors would, theoretically, mean that smaller leak/burst events could be localised. However, other sources of uncertainty such as the customer demands and the corresponding effect this has on the pressure, can easily overshadow the uncertainty of the instrument. Therefore, with the current state of WDS, where widespread monitoring of customers is not commonplace it is envisaged that increasing the accuracy of pressure instruments used would not have a meaningful impact upon the minimum size of leak/burst events which can be localised.

In relation to the differences between the hydraulic model and the real DMAs which have been considered in this thesis, the changes in pressure are typically larger in the real DMAs. The hydraulic model which was used to model the engineered events (which is the same as the model used by the sensor placement technique) was not highly calibrated, making differences very likely. The leak/burst localisation technique is able to determine the same search area even when differences exist between the real and modelled data as was demonstrated in Figure 6-15 in section 6.3.2.6. This was also demonstrated for engineered events in DMA 123-09 for which the two search areas agreed perfectly even though there were differences in the changes in pressure between the model and the real DMA. An example is provided in Figures 7-1 and 7-2 to further demonstrate this. This fact does not impact on the performance of the leak/burst localisation technique because the size of the event is not considered by the leak/burst localisation technique. It does mean that the modelled leak/burst event scenarios are conservative and that even smaller events in real DMAs can be localised than is predicted by the hydraulic model. However, the optimal sensor configurations will not have considered the smaller event sizes for a DMA where the real changes in pressure are larger than those produced by the model.

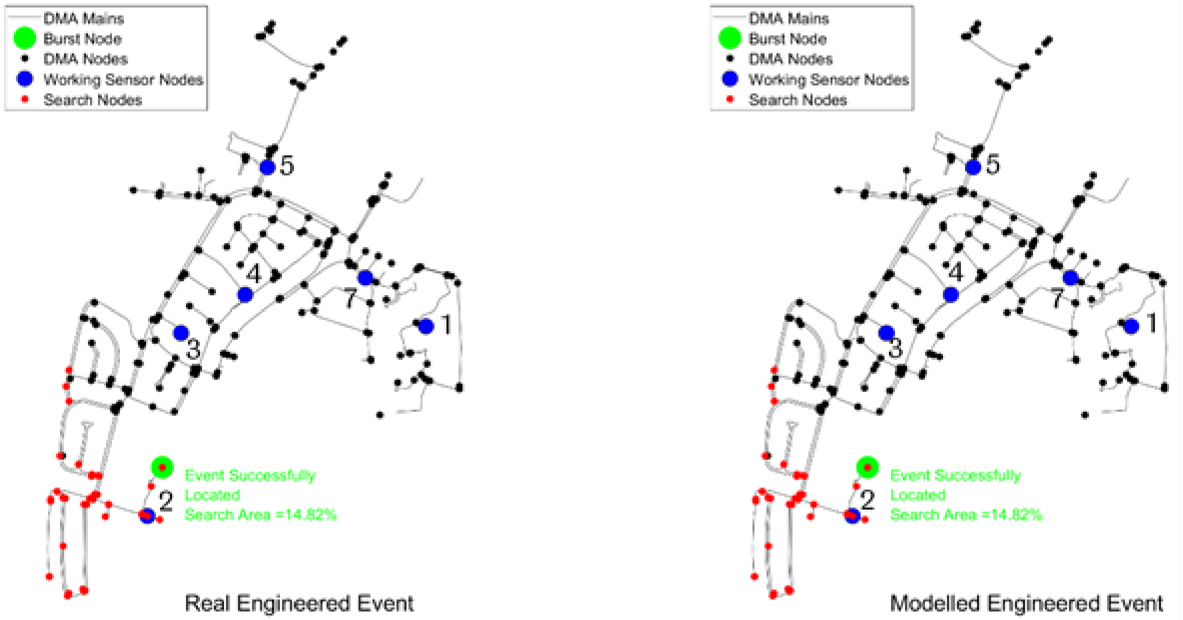


Figure 7-1 *Search area determined for a real and modelled engineered event in DMA 123-09*

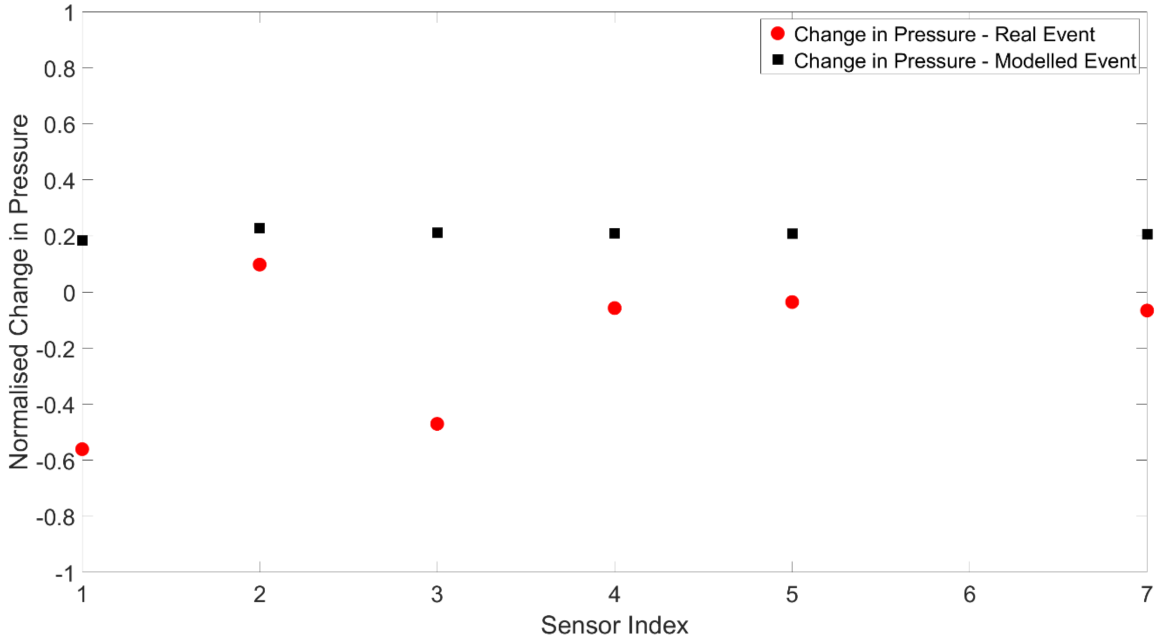


Figure 7-2 *Changes in pressure used to determine the search areas for the real and modelled engineered event in Figure 7-1*

A related issue pertinent to the sensor placement technique is that the grouping of leak/burst events relies heavily on this the accuracy of pressure sensors and more accurate sensors, which can better distinguish between different leak/burst event locations in the hydraulic model would tend to produce higher numbers of groups This has a practical implication for the time taken to determine the optimal sensor configurations although this is of less concern than the issues which have been raised previously in this section.

Using a data-driven leak/burst localisation technique means that the uncertainties in the real WDS will inevitably influence the localisation performance. This is related to sensor uncertainty in that the actual pressure at a point in a real WDS will not be perfectly reflected in the measured pressure registered by an instrument. When comparing a real network with a model the biggest (but not only) source of uncertainty are the customer demands. Customer demands are not usually known because metering every customer is still not feasible (though roll out of AMR will improve this in the future) for water companies for reasons of cost. The leak/burst localisation technique deals with the problem of uncertainty by defining a search area rather than aiming to identify a single leak/burst event location such as the techniques proposed by Casillas et al. (2015) and Perez et al. (2011). Defining a search area gives a larger area to search, which is a drawback, but the advantage is that correctly localising leak/burst events is easier. This helps to overcome the problem of uncertainty because a larger apparent change in one or more of the pressure measurements is required to significantly change where the search area is determined in the DMA. The use of the SC-IDW interpolation technique is very beneficial in this respect because for every location in the DMA all pressure sensor measurements are considered to estimate the change in pressure. Therefore, the method is less reliant on individual pressure measurements to correctly determine the search area. Combining this with the optimal parameters, which trade off the size of the search area with the number of correctly localised leak/burst events means that the smallest search areas are produced which consistently localise many leak/burst events correctly. Using the night-time pressure, when the customer demand is minimised, means that the uncertainty associated with the customer demands is also minimised. This means that variations in pressure which arise as a result of the variation in customer demand between the night containing the leak/burst event and the previous days which represent the normal state of the DMA without any leak/burst event present. Some variation in the customer demands will still be present, even when no leak/burst events are present, which is why the average of 14 days is taken to represent that normal state of the DMA. Having said this, the leak/burst localisation technique will still face difficulties if a significant difference in the demand is present on the night containing the leak/burst event as there is no way to distinguish between the change in pressure due to the leak/burst event and due to the change in customer demand.

By performing the localisation during the night, the maximum time that can be taken to localise a leak/burst event is 24 hours, assuming that the leak/burst event starts just after the end of the analysis window of the previous night. The current situation whereby no automated analysis of data from DMAs is conducted can be vastly improved even by using a daily leak/burst localisation technique in concert with a near real time leak/burst detection system. Even the most advanced water companies (who perform automated detection) still do not perform the localisation step using additional data from their DMAs.

7.2.5. Automatic determination and grouping of leak/burst event scenarios for sensor placement

A compelling issue for any sensor placement technique is to determine what leak/burst event scenarios should be considered by the sensor placement technique. The most typical approach is to select a single size of leak/burst events (specified using the emitter coefficient) and model this size of event at many locations throughout a DMA to produce a sensitivity matrix (Farley et al., 2008, Perez et al., 2009). The drawback of this approach is that the sensor configurations will only be optimal with respect to the leak/burst event scenarios which are considered by the sensor placement. The main benefit is that the number of leak/burst event scenarios is the same as the number of possible leak/burst locations. There exists a trade-off between how many scenarios must be considered (with implications for the computational time taken to run the sensor placement technique) and how specific the optimal sensor configurations are in terms of the range of leak/burst events which can arise in a DMA. Techniques exist which consider multiple leak/burst event sizes (Casillas et al., 2013, 2015, Cuguero-Escofet et al., 2015) however in these cases the leak/burst event sizes are specified manually and do not consider the changes in pressure which occur as a result of the selected leak/burst event sizes. The range of leak/burst event sizes is customised for each DMA and is determined automatically by considering the changes in pressure and flow which occur for each leak/burst event size which is a significant benefit over selecting the sizes without reference to the hydraulic changes occurring as a result. This is the first automatic procedure for determining a range of leak/burst event sizes for sensor placement in a DMA for the purpose of leak/burst localisation.

Another novel procedure, developed as part of the sensor placement technique, groups together the leak/burst event scenarios to reduce the number of leak/burst events which must be considered by the sensor placement technique. The purpose of this is to reduce the computational time taken to determine the optimal sensor configurations. The criteria for grouping events together was whether there was a significant difference in the changes in pressure caused by them when compared to the accuracy of the pressure instruments. A previous attempt at reducing the size of the search space by Sophocleous et al. (2019) grouped together the leak/burst events by proximity. However, even leak/burst events which are close together can have very different effects on a DMA. Using the effect on pressure directly to group the leak/burst events overcomes this problem. The procedure for grouping leak/burst events is not necessary when the number of leak/burst event scenarios is small or when computationally simple techniques are being used. In future it is envisaged that the computational complexity of leak/burst localisation techniques will increase and the grouping procedure presented here provides a way of using these techniques in conjunction with a GA whilst not incurring prohibitive run times.

Case study 5.8 in section 5.4.5 demonstrated how the parameters of the technique for automatically determining the leak/burst event sizes affected whether a valid range of leak/burst event sizes could be produced for a DMA. The event size increment determines the resolution with which the leak/burst event sizes are determined and using too high a value for this parameter negatively impacted the accuracy of the maximum and minimum leak/burst event sizes. Increasing the maximum allowable increase in flow could overcome the problem of the lack of accuracy in determining the leak/burst event sizes although this then means that larger leak/burst events, which would typically be reported would then be considered by the sensor placement technique. Case study 5.8 also outlined a procedure for determining whether a DMA can be considered as suitably sensitive to leak/burst events to be suitable for deploying pressure sensors. If a DMA only produces valid results, for multiple different event size increments, when a high value for the maximum allowable increase in flow was specified then this indicates that it would not be a suitable candidate for deploying pressure sensors for localising unreported leak/burst event sizes. In this context the acceptable value for the maximum allowable increase in flow can be decided individually for a DMA or specified across an entire WDS, in line with the regulatory targets of a water company.

7.2.6. Sensor locations are similar to those which are used for calibration potentially enabling multipurpose sensor networks

By selecting instrument locations which may also be good for calibration and reporting purposes some of the marginal cost can be negated as other sensors would be needed for these purposes anyway. This point was stated by Hart and Murray (2010) in relation to sensor placement for contamination detection. The sensor placement technique used by the framework of methods and those used for hydraulic model calibration are often aiming to identify locations for pressure sensors which are highly sensitive to changes in flow. The sensor placement technique used by the framework of methods tends to place sensors at the edges of a DMA or at the end of pipes within dendritic sections of the DMA. This is where sensor placement techniques for hydraulic model calibration (see Kapelan et al., 2005) also favour placing sensors although they will not be optimally placed for this purpose. From a water company perspective having sensors permanently deployed within a DMA removes the need to temporarily deploy pressure sensors prior to performing flushing events to allow hydraulic model calibration. This would reduce the amount of time spent by operatives in the field for this purpose which is an additional benefit.

7.3. Technical Discussion

7.3.1. Use of inverse-distance weighting for computational simplicity and efficiency

A fundamental part of the leak/burst localisation technique is the SC-IDW technique which is used to estimate the change in pressure throughout a DMA from a limited number of pressure measurements. The SC-IDW technique is an improved version of IDW interpolation which accounts for the layout of pipes in a DMA. Using SC-IDW versus the alternative spatial/geostatistical techniques presents several advantages. The first key advantage is that IDW is simpler and therefore takes significantly less time to compute than, for example, OK (Soldevila et al., 2019). This is primarily because for IDW only one parameter must be specified and finding the best value for this parameter can be achieved by the sensor placement technique alongside the sensor configurations. For OK there are a several model parameters which must be specified (type of variogram model, variogram nugget, variogram sill and variogram range – see Cressie, 1993) and they are usually fitted empirically using some error minimisation scheme. For each leak/burst event this means that the best value for each of the parameters must be found by trialling a large number of combinations of all parameters because the model parameters are derived from the data itself. A further point to this is that the accuracy of OK (as well as other Kriging variants) is very closely linked to how well the variogram model fits the data. A poorly fitted variogram does not preclude the use of OK but the status of OK as being more accurate than other techniques only holds when the variogram fits the data well (Cressie, 1993). Another interesting point is that because a different model must be used for each data set to ensure a good variogram model fit, the same model cannot be used for a real event as would be used for the same modelled event. This means that optimising the variogram model parameters within the sensor placement technique would produce an infeasibly large set of decision variables and search space. The computational time is critical where the unbinarised sensitivity matrix is being analysed by a secondary statistical/analytical technique, such as SC-IDW. For techniques using the binarised sensitivity matrix the computational task is normally much less burdensome but cannot be used with spatial analysis techniques which require continuous rather than discrete data to produce meaningful results.

In terms of the performance when using different spatial/geostatistical techniques a summary of the achieved leak/burst localisation performance was given in section 7.2.4 of this chapter. The main point was that comparable performance was achieved, for similar size of engineered event, by using the SC-IDW technique and using OK with the technique developed by Soldevila et al., (2019) which indicates that using the more complex geostatistical technique did not provide additional performance benefits.

7.3.2. Number of sensors

The number of additional pressure sensors deployed throughout a DMA is a critical factor which dictates the achievable leak/burst localisation performance. Several of the engineered events, for which the localisation results were given for the optimal sensor configurations in Table 6-3 in section 6.3.2.4, showed that increasing the number of sensors led to smaller search areas in the case of DMA 304-03. Considering the sensor placement performance, the same trend was seen when the average performance across hundreds of leak/burst events was determined. Critically, the improvement in performance diminished as higher numbers of sensors are considered meaning that water companies would likely prefer to deploy fewer sensors although this would depend upon their strategic aims. The current rate of pressure sensor deployment in WDSs, which is mostly not for the purpose of leak/burst localisation but for ensuring that adequate pressure is supplied to customer by measuring the pressure at a critical point or several critical points, is typically only one or two instruments per DMA. Given this context, a conservative strategy would favour placing a small number of additional optimal pressure sensors. It is worth highlighting that this number does not need to be fixed and the density of sensors could be increased over time by adding to previously deployed sensors and moving existing sensors so that the optimal configuration is always used.

The sensor placement results for DMA 304-04 demonstrated that no improvement in performance was seen for numbers of sensors above 3. In this case it would mean that, by default, 3 sensors would be chosen. This is automatically determined by the GALAXY MOEA such that if no improvement in performance in seen above a certain number of sensors then the results produced will not consider the numbers of sensors for which no performance benefit is seen. This also demonstrates that the constraint placed upon the sensor placement technique of choosing between 3 and 10 sensors was valid because in every case examined it was clear that less than 10 sensors would be deployed in practise due to diminishing return of increasing numbers of sensors much higher than 5. This also provides a useful "stepping stone" as it is higher than the number of sensor currently deployed and is likely to be less than the number deployed in the near future as the cost of sensors continues to fall, particularly in light of increasing regulation and stricter performance targets with more severe penalties related to the effects of leak/burst events including water supply interruptions and leakage which are common performance commitments for all water companies in the UK for the current AMP cycle (PR19).

7.3.3. Reduced parameterisation of the GALAXY MOEA compared to other genetic/evolutionary algorithms

It is well known that a major difficulty faced when using GAs/evolutionary algorithms is that many parameters often need to be specified on a case specific basis to ensure good performance. Only two parameters, namely the population size and the number of function evaluations, need to be specified in order to run the GALAXY MOEA. This fact was demonstrated by the developers of the GALAXY MOEA (Wang et al., 2017) by testing its performance on a wide range of benchmark optimisation problems from the field of WDS analysis. Another parameter can be specified which is the range of mutation associated with the dither creeping search operator. Wang et al. (2017) stated that the value of this parameter affects the likelihood that a decision variable is modified in a particular direction. The default value of 0.7 for the range of mutation for the dither creeping search operator, as used by Wang et al. (2017) was used by the sensor placement technique. Wang et al. (2017) also stated that the performance of the GALAXY MOEA was not found to vary significantly for different values of this parameter.

In the context of a water company, where expertise in techniques such as the GALAXY MOEA is not widely available, reducing the parameterisation down to just selecting the population size and the number of function evaluations greatly simplifies its implementation.

7.4. Practical Issues

7.4.1. Real world considerations are accounted for by the sensor placement objective function

As part of the sensor placement technique, events with multiple search areas were penalised. This was included in the sensor placement objective function to dissuade the sensor placement technique from selecting sensor configurations for which large numbers of leak/burst events produced multiple search areas. This was based on the consideration that travelling between multiple search areas in a DMA would increase the time taken to localise a leak/burst event. The effect of the penalty was demonstrated in case study 5.6 in section 5.4.3. The penalty significantly reduced the incidence of multiple search areas being produced but this came at the cost of a modest increase in the average size of the search area. The size of the search area was measured using pipe length so that the localisation performance reflected the reality of localising leak/burst events in WDS. Using a distance-based measure of leak/burst localisation accuracy is not new within the literature. For example, Soldevila et al. (2019) used the distance between the actual leak/burst event location and the location identified by the leak/burst localisation technique. However, this metric assumes that the field teams would travel directly from the suspected location to the actual location which is not realistic. Field teams would most likely radiate out in multiple directions from the suspected location which is why the total length of pipes falling in an area is a more suitable metric for leak/burst localisation performance.

The potential sensor locations were limited to hydrants only. The leak/burst performance reduction was found to be between 2-3% when compared to considering all junctions as candidate sensor locations which is not a significant reduction. This is even less significant given that installing sensors at hydrants instead of on a pipe costs much less and is deemed favourable considering the trade-off between the cost and the leak/burst localisation performance.

7.4.2. Impacts upon leak/burst intervention

The main benefit of the framework of methods presented in this thesis is to reduce the amount of time taken to find the precise location of leak/burst events within WDS by providing an approximate area within the DMA to be searched by field teams. By detecting a leak/burst event and approximating its size (which is achieved using an event recognition system used by United Utilities) and then determining a search area the total cost of finding and repairing it can be minimised by shortening its lifecycle. This being said not all leak/burst events need to be repaired until they present a significant enough risk to a water company due because in some cases the cost of repairing the leak/burst event is higher than the cost of letting it continue. If the size of the event, its effect on the pressure in a DMA and the amount of the DMA which would need to be searched in order to find it are known then an approximate cost of repairing the leak/burst could be inferred. In this context, an informed decision about whether to find and fix the leak/burst events can be taken by operators which will lead to improved management of new leak/burst events in WDSs.

There is the potential that if a leak/burst event is not fixed when first detected that it "breaks out" into a larger event. In this case the differential changes in pressure could increase meaning that the search area size may be reduced. This could mean that, in some cases, monitoring known leak/burst events over time and fixing them when they become sufficiently problematic would represent the most efficient approach rather than attempting to fix every event as soon as it arises and its approximate location can be determined.

7.5. Summary of Discussion

Presented in this chapter has been a discussion of the most pertinent issues relating to the performance and applicability to the water industry of the framework of methods. Several key issues were identified which are summarised in the bullet points below:

- The framework is integrated so that the leak/burst localisation technique is used by the sensor placement technique to determine the optimal configuration(s) of pressure sensors. This ensures that the sensor configurations consider not only the hydraulic changes which are caused by leak/burst events but also the technique which is being used to localise them. A range of optimal sensor configurations are determined, although this is not guaranteed for every DMA, so that they can be presented to decision makers to make an informed decision considering the trade-off between the number of sensors and the leak/burst localisation accuracy.
- The leak/burst localisation technique is data-driven and does not require a hydraulic model to determine the approximate location of leak/burst events. The past pressures determined for each pressure sensor in the DMA, calculated over 14 nights, is compared to the pressure in the presence of a new leak/burst event. Using the night means that the uncertainty in the customer demands is minimised and has less influence over the leak/burst localisation performance.
- Leak/burst events as small as 3.5% of the peak daily flow were correctly localised using as few as 3 optimal pressure sensors. This was achieved even though the size of the engineered events was smaller than the minimum theoretical leak/burst event sizes determined prior to running the sensor placement technique.
- The sensor placement technique does not rely on a well-calibrated hydraulic model because the leak/burst localisation technique uses the differential changes in pressure throughout the DMA to infer the approximate location of a leak/burst event. This is also aided by the use of the SC-IDW technique because, for every location in a DMA, multiple pressure sensors are used to estimate the change in pressure. This reduces the reliance on a single instrument which, in

turn, reduces the influence of sensor uncertainty on the determined search area.

Further to the key items listed above some relevant technical details of the framework of methods and their implications on the applicability of the framework of methods within the water industry were also identified. The discussion presented here demonstrates that the research aims and objectives which were stated in Chapter 3 have been achieved.

Chapter 8 - Conclusions and Future Work

8.1. Introduction

A novel, integrated optimal sensor placement and leak/burst localisation framework has been developed, tested and validated which (i) allows the Pareto front of optimal sensor configurations, in terms of both number and location of pressure sensors, to be determined for a DMA and (ii) analyses the night-time pressure data collected from those sensors to determine the approximate location of new leaks/bursts in that DMA. This can be passed to field teams in order to reduce the time taken to quickly effect repairs, which in turn, can enable water companies to mitigate the multiple potential negative consequences associated with them.

In the remaining sections the general conclusions drawn as a result of the work presented in Chapters 2-7 are given in order of their significance. Then specific conclusions related to each of the objectives given in Chapter 3 are given.

8.2. General Conclusions

- By incorporating the novel sensor placement technique into the integrated framework (and using the same SC-IDW interpolation technique at the heart of both steps) the leak/burst localisation accuracy can be improved. This was demonstrated by comparing the localisation performance of the integrated framework against a baseline sensor placement technique. Validation demonstrated that the novel sensor placement technique (as part of the integrated framework) can tolerate inconsistencies/inaccuracies which typically exist in the water company hydraulic models which have been used to determine the optimal sensor configurations. The extent to which improving model calibration could improve the leak/burst localisation performance of the integrated framework was not investigated due to the increased labour and cost associated with this.

- The leak/burst localisation technique can be run by control centre staff following receipt of event alarms raised by a leak detection technique. This will allow the determined approximate location of a leak/burst event to guide field teams to the approximate location of the event to reduce the time taken to locate it with hardware-based techniques (i.e. noise-correlators). This can enable repairs to be promptly carried out to minimise disruption to customers by, for example, minimising the number of customers without supply during waking hours and which can lead to better performance in several regulatory targets including water supply interruptions and leakage as set out under the common performance commitments set by Ofwat for the 2019 price review.
- The results of validation of the integrated framework using engineered leak/burst events, by opening fire hydrants, in a real DMA in United Utilities WDS demonstrate that it can typically locate leaks/bursts to a small fraction of the DMA (the best localisation performance achieved was approximately 14%). The integrated framework has been proven to be able to successfully locate leaks/bursts as small as 3.5% of the peak daily flow with as few as 3 additional pressure sensors installed.

8.3. Objective Specific Conclusions

The specific conclusions for each objective are presented in the same order as the objectives were given in Chapter 3.

8.3.1. Objective 1: Performance Metrics

- Three performance metrics were identified to allow the performance of the sensor placement technique and the leak/burst localisation technique to be determined. The average size of the search area, the number of leak/burst events correctly localised and the number of leak/burst events for which multiple search areas are produced were used during all applications of the framework of methods presented in this thesis. The same set of performance metrics were used by both the sensor placement technique and the leak/burst localisation technique to ensure consistency. The sensor placement technique can identify the sensor configurations and combination of parameters which gives the best performance considering all performance metrics simultaneously.

- The performance objectives consider the accuracy of leak/burst localisation by minimising the size of the search area as much as possible whilst also ensuring that the leak/burst event location is contained within the search area for the maximum number of leak/burst events.

8.3.2. Objective 2: Constraints

- The sensor locations were constrained to being installed at hydrants due to the ease of installation and reduced cost when compared to digging down to a pipe and installing a sensor within a chamber. Although leading to a small loss in leak/burst localisation performance it was determined that the loss of performance was minimal in comparison with the reduced cost of installing sensors at hydrants. Given that it is common for there to be tens of hydrants per DMA and that the high cost of building chambers is an issue throughout all WDSs, deploying sensors at hydrants is the suggested approach for all DMAs.
- For each DMA considered in this research the likely number of sensors to be deployed would be less than 5 by considering the trade-off between the achievable leak/burst localisation performance contained in the Pareto fronts and the number of deployed sensors. This demonstrated that limiting the maximum number of sensors to 10 within the sensor placement technique was valid for the developed sensor placement technique. However, whether this is true for DMAs which have not been investigated in any of the presented case studies would need to be determined for each DMA. This cannot be generalised to all DMAs based upon the limited number of DMAs which have been used to perform the case studies presented in this thesis.

8.3.3. Objective 3: Leak/burst event size and changes in pressure

- Leak/burst event sizes which are typically not reported, given the current state of WDS monitoring, are targeted by the framework of methods. The size of leak/burst events which caused a significant change in pressure varies throughout a DMA depended upon the hydraulic conditions, driven by the physical characteristics of the pipes and the customer demand, throughout the DMA. This means that the smallest leak/burst event size is specific to each leak/burst event location and must be determined for each separately.
- The pressure was most sensitive to leak/burst events during the night because the leak/burst event flow, for a given emitter coefficient, was maximised in relation to the customer demand. Given the inherent uncertainty associated with the customer demands, which is largest during the day, the change in pressure for unreported leak/burst events was masked during the day but was readily apparent during the night.
- An important consideration, which is uniquely accounted for in the framework of methods, is that the accuracy of the pressure instruments is used to determine the smallest size of leak/burst events for the sensor placement technique to consider. The accuracy of the instruments limits the minimum change in pressure which can be measured with certainty and using an instrument with lower accuracy limits the smallest size of leak/burst events which can be considered.
- Using the differential changes in pressure, by comparing the measured changes in pressure for multiple pressure sensors, emphasised the local effects of a leak/burst event which could be used to infer its approximate location.

8.3.4. Objective 4: Automatic leak/burst event sizes and grouping of leak/burst event locations

- A novel procedure for determining the smallest and largest leak/burst event size for all considered leak/burst event locations was developed. The procedure is automatic and requires the specification of three parameters so that all leak/burst event sizes can be determined for an entire DMA. This precludes the need for a detailed investigation or sensitivity analysis of the leak/burst event sizes which is typically conducted for this purpose.

- The accuracy of the pressure instruments, which in this thesis was considered to be 0.1m due to resolution of the pressure data files supplied by the instrument manufacturers, was used to calculate the smallest emitter coefficient which caused a large enough change in pressure at any single point in the DMA assuming that the DMA was monitored for pressure at all candidate sensor locations. This ensures that events which are smaller than the minimum size which can be measured are not considered by the sensor placement technique.
- The maximum size of events was limited to those falling below a user specified maximum allowable value. In this thesis 10% of the peak daily flow was used because in many cases events larger than this are more readily apparent and can be localised using other means. This ensures that the optimal sensor configurations are targeted specifically to leak/burst event sizes which are not otherwise readily apparent. Specifying the maximum size of leak/burst events in this way means that the sensor placement technique did not consider events which would typically be reported due to the widespread and large changes in pressure which they often cause in a DMA.
- A novel leak/burst event grouping procedure was used to group together leak/burst event locations using only the changes in pressure which were caused by them. This significantly reduced the required computational effort and ensured that leak/burst events which caused very similar changes in pressure for all candidate sensor locations were not considered separately by the sensor placement technique. The accuracy of the pressure instruments was used so that if the changes in pressure for multiple leak/burst events were did not differ significantly enough to be measured by any configuration of sensors they were treated as part of the same group of leak/burst events.

8.3.5. Objective 5: Develop a leak/burst localisation technique

- A data-driven technique for determining the change in pressure, as a result of a new leak/bust event was developed which compares that past behaviour of a DMA which its behaviour after the start of a new leak/burst event. Many model-based leak/burst localisation techniques compare the model data directly to the real data to determine the change in pressure and, therefore, differences between the model and the real DMA can hamper their performance. This means that the hydraulic model must be highly calibrated for these techniques to perform well. Crucially, the leak/burst localisation technique presented in this thesis does not require a hydraulic model to be compared with the measured pressures from a real DMA which reduced its reliance on a highly calibrated hydraulic model when compared to the model-based leak/burst localisation techniques.
- A technique for normalising the changes in pressure relative to average change in pressure for multiple leak/burst events was developed to overcome the differences in sensitivity due to geometric effects. This helped to overcome difficulties of localising leak/burst events near to the inlet of a DMA which can cause larger changes in pressure at downstream sensor locations which are not proximate to the leak/burst event location.
- The leak/burst localisation technique was developed and tested using engineered event data collected from a real DMA. The number of previous days used to infer the approximate location of the engineered events was investigated and using the average of 14 previous days demonstrated the performance, in terms of the number of events correctly localised. Using 14 previous days reduced the reliance on any single day which can overcome the uncertainty in the customer demand during the night whilst adequately excluding the long-term variations in pressure which can arise.

8.3.6. Objective 6: Develop a spatial/geostatistical technique

- A novel SC-IDW interpolation technique was developed as part of the leak/burst localisation technique. The SC-IDW interpolation technique combines the changes in pressure from multiple pressure sensors to build a picture of the change in pressure for the entire DMA. The estimated change in pressure for each location is determined using the measured changes in pressure from all pressure sensors. This means that the reliance on any single pressure sensor is reduced. This characteristic means that the leak/burst localisation technique is less sensitive to the failure of a single sensor.

- The use of IDW as opposed to more complicated and computationally expensive techniques (such as OK or OC) means that the optimal parameters can be determined concurrently with the optimal sensor configurations to ensure that the accuracy of the SC-IDW interpolation technique is maximised.
- The distance used by the SC-IDW interpolation technique is determined as the shortest path between two points in a DMA travelling along the pipes. This ensures that the SC-IDW interpolation technique respects the topology of the pipes in the DMA unlike using Euclidean distance as per traditional IDW interpolation. This distance function improved the differentiation between pipes which are close together but not connected to each other to determine the search area more accurately.

8.3.7. Objective 7: Develop optimal sensor placement technique

- A sensor placement technique was developed in conjunction with the leak/burst localisation technique to determine a Pareto front of optimal sensor configurations (and combinations of parameters) with varying numbers of sensors. The GALAXY MOEA was used, by developing a new objective function, to determine the sensor configurations and combinations of parameters which simultaneously minimise the required number of sensors and maximise the leak/burst localisation performance relative to the selected performance metrics.
- The objective function for the GALAXY MOEA, presented in this thesis, uses the localisation accuracy directly which ensures that the sensor configurations which find the smallest search areas (and therefore the greatest benefit) are identified. By formulating the sensor placement problem as a multi-objective

optimisation problem, a range of optimal solutions with varying numbers of sensors and differing degrees of leak/burst localisation performance are presented, so that the preferred solution can be chosen.

- By selecting the GALAXY MOEA to determine the optimal sensor configurations alongside the combinations of leak/burst localisation parameters the issue of parameterisation of the MOEA is greatly reduced. The GALAXY MOEA only requires two parameters, namely the size of the population and the number of function evaluations, to be specified prior to running the sensor placement technique.

8.3.8. Objective 8: Field test/validate the framework of methods

- The framework of methods was validated using two real DMAs in United Utilities WDS. A total of 20 engineered events were conducted in the two DMAs and 29 pressure sensors, collecting a measurement every minute, were installed prior to the start of the engineered events. A flow rate of approximately 0.6l/s was used for all engineered events which was smaller than 5% of the peak daily flow for both validation DMAs. The constrained optimal sensor configurations were determined from the deployed sensor locations by using a heavily constrained version of the sensor placement technique. Baseline sensor configurations were also selected at random from the deployed sensor locations and a comparison between the optimal and baseline configurations was conducted. The constrained optimal sensor configuration successfully localised engineered events as small as 3.5% of the peak daily flow using as few as 3 pressure sensors. The best performance was achieved for engineered event 8 where the search area, using 5 sensors, was 19% of the DMA.
- For both validation DMAs the average size of the search area for the constrained optimal sensor configurations was smaller than for the baseline configurations whilst also localising more leak/burst events correctly. The difference in performance was not significant due to the highly constrained nature of the sensor placement technique, which was used out of necessity due to limitations in the available sensor locations, as a result of the preselected sensor

deployment. Another contributing factor to the very similar performance between the constrained optimal and baseline sensor configurations was the small size of the engineered events which were used.

- The engineered event scenarios used to validate the framework of methods were more difficult to localise than the leak/burst event scenarios generated by the sensor placement technique. This was due to a combination of the small size of the engineered events and the locations which were selected.

8.4. Recommendations for future work

In this section some areas for future research have been identified which are listed in the bullet points below.

- Investigate how the level of hydraulic model calibration affects the performance of the sensor placement technique. Although the leak/burst localisation technique presented in this thesis does not require a highly calibrated hydraulic model in order to successfully and accurately localise leak/burst events, the sensor configurations are dependent to some extent upon the agreement between the hydraulic model and the real DMA. An interesting avenue of further research is to determine the relationship between the level of hydraulic model calibration and the sensor configurations which are determined as a result. Further to this the leak/burst localisation performance, determined using engineered events, should be ascertained for the optimal sensor configurations with different levels of hydraulic model calibration.
- Perform a detailed cost-benefit analysis for deploying additional pressure sensors for leak/burst localisation and build this into the sensor placement technique. There is a cost which can be associated with each type of individual leak/burst event which can occur in a DMA which is related to regulatory targets (and associated penalties) and the cost of repairing it. For each DMA, an associated cost of leak/burst events can be determined and compared to the cost of deploying different numbers of pressure sensors. The cost of deploying pressure sensors was assumed as directly proportional to the number of

sensors in a DMA but in reality the cost will not be linear due to economies of scale.

- Develop a methodology for prioritising which DMAs should be targeted for deploying pressure sensors. It is envisaged that a phased approach to deploying pressure sensors would be taken to minimise the required amount of capital investment. Therefore, knowing which DMAs are good candidates for deploying pressure sensors by considering the historical leak/burst performance of each DMA and the potential sensor placement performance would ensure that the most beneficial DMAs can be targeted.
- Accounting for the typical variations in demand which occur in a DMA and incorporating this into the leak/burst localisation technique. A limiting factor of the smallest size of the leak/burst events which can be localised (and how accurately they can be localised) is that smaller events can be more easily masked by normal variations in the demand and the resulting affect this has on the pressure throughout a DMA. Adapting techniques for stochastically estimating the demand (Creaco et al., 2017) to better distinguish between the changes in pressure due to the leak/burst and the change in pressure due to an increase in demand could improve the performance of the leak/burst localisation technique.
- Investigate the potential for dual-use sensor networks by measuring multiple parameters (including pressure, flow, turbidity and water quality at the same point) and developing a technique for determining the optimal locations for measuring these parameters. There are other undesirable phenomena in WDSs that can be aided by continuous monitoring including discolouration and low pressure/loss of supply. Further development and customisation of the proposed framework of methods to other types of events, using different types of sensor for which the optimal locations are selected, could yield performance benefits in these areas. Further to this, the use of continuous monitoring could enable better visibility of a WDS to improve other water company functions such as better hydraulic model maintenance and reporting to Ofwat.

To summarise, improving the management of WDSs in relation to their performance and efficiency, is a compelling issue for water companies around the world. Increased monitoring of WDSs allows greater visibility of WDS performance so that timely interventions can be made to maximise customer service and ensure that the water supplies are secure for all customers in the future. Continuous pressure monitoring of WDSs, using multiple optimally placed pressure sensors per DMA, can enhance the leak/burst localisation process so that water companies can quickly react to new leak/burst events to mitigate the negative impacts they cause.

Bibliography

Abadie, J. (1970). Application of the GRG method to optimal control problems. In: Integer and Nonlinear Programming (J. Abadie, ed.). North-Holland Publishing Company, Amsterdam, North Holland, The Netherlands.

Affinity Water. (2018). Our Business Plan for 2020-2025: Appendix 9 - Ensuring Long Term Resilience. Available at: https://www.affinitywater.co.uk/docs/corporate/plans/appendix-9-ensuring-long-term-resilience.pdf (Accessed 09th April 2021).

Aksela, K., Aksela, M. and Vahala, R. (2009). Leakage Detection in a Real Distribution Network Using a SOM. Urban Water J. 6(4):279-89.

Alegre, H. (2007). Performance Indicators for Water Supply Services. 2. ed., Reprinted. Manual of Best Practice. London, IWA.

Alexander, C. K. and Sadiku, M. N. O. (2013). Fundamentals of Electric Circuits. 5th ed. New York: McGraw-Hill.

American Water Works Association (AWWA). (2003). Water Audits and Loss Control Programmes. 4th ed.

Andreou, S. A., Marks, D. H. and Clark, R. M. (1987). A New Methodology for Modelling Break Failure Patterns in Deteriorating Water Distribution Systems: Theory. Adv. Water Resources. 10(1):2-10.

Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys. 4:40-79.

Atef, A., Zayed, T., Hawari., A,. Khader, M. and Moselhi, O. (2016). Multi-Tier Method Using Infrared Photography and GPR to Detect and Locate Water Leaks. Autom. Constr. 61:162-70.

Bakker, M., Vreeburg, J. H. G., Van Schagen, K. M. and Rietveld, L.C. (2013). A Fully Adaptive Forecasting Model for Short-Term Drinking Water Demand. Environ. Modell. Softw. 48:141–51.

Bakker, M., Vreeburg, J. H. G., Van De Roer, M. and Rietveld, L. C. (2014). Heuristic Burst Detection Method Using Flow and Pressure Measurements. J Hydroinform. 16(5):1194-1209.

Beck, S. B. M., Curren, M. D., Sims, N. D. and Stanway, R. (2005). Pipeline network features and leak detection by cross-correlation analysis of reflected waves. J Hydraul. Eng. 131(8):715-23.

Bell, E. T. (1934). Exponential Numbers. Amer. Math. Monthly 41:411-419.

Bishop, C. M. (1995). Neural Networks for Pattern Recognition. Oxford University Press.

Blesa, J., Nejjari, F. and Sarrate, R., 2016. Robust sensor placement for leak location: Analysis and design. J Hydroinform. 18(1):136-148.

Boccelli, D. L., Tryby, M. E., Uber, J. G., & Summers, R. S. (2003). A reactive species model for chlorine decay and THM formation under rechlorination conditions. Water Res. 37(11):2654–2666.

Borovik, I., Ulanicki, B. and Skworcow, P. (2009). Bursts identification in water distribution systems. In Proceedings of World Environmental and Water Resources Congress, Kansas City, Missouri.

Boulos, P.F. and Wood, D.J. (1990). Explicit calculation of pipe-network parameters. J Hydraul. Eng. 116(11):1329-44.

Boulos, P.F., Karney, B.W., Wood, D.J. and Lingireddy, S. (2005). Hydraulic transient guidelines for protecting water distribution systems. J Am. Water Works Ass. 97(5):111-24.

Boulos, P.F. and Aboujaoude, A.S. (2011). Managing leaks using flow step-testing, network modeling, and field measurement. J Am. Water Works Ass. 103(2):90-7.

Brunone, B. (1999). Transient Test-Based Technique for Leak Detection in Outfall Pipes. J Water Res. Plan. Man. 125(5):302-6.

Bush, C. A., & Uber, J. G. (1998). Sampling Design Methods for Water Distribution Model Calibration. J Water Res. Plan. Man. 124(6):334–44.

Caldecott, R., Poirier, M., Scofea, D., Svoboda, D.E. and Terzuoli, A.J. (1988). Underground Mapping of Utility Lines Using Impulse Radar. In Proceedings F: Communications, Radar and Signal Processing. 135(4):343-53.

Candelieri, A., Soldi, D., Conti, D., & Archetti, F. (2014). Analytical Leakages Localization in Water Distribution Networks through Spectral Clustering and Support

Vector MACHINES. In Proceedings of the 16th International Conference on Water Distribution Systems Analysis, Bari, Italy.

Caputo, A. C. and Pelagagge, P. M. (2002). An inverse approach for piping networks monitoring. J Loss Prevent. Proc. 15(6):497-505.

Caputo, A.C. and Pelagagge, P.M. (2003). Using neural networks to monitor piping systems. Process Saf. Prog. 22(2):119-28.

Casillas, M.V., Puig, V., Garza-Castanon, L.E. and Rosich, A. (2013). Optimal sensor placement for leak location in water distribution networks using genetic algorithms. Sensors. 13(11):14984-15005.

Casillas, M.V., Garza-Castanon, L.E. and Puig, V. (2015). Optimal sensor placement for leak location in water distribution networks using evolutionary algorithms. Water. 7(11):6496-6515.

Christodoulou, S. E., Gagatsis, A., Xanthos, S., Kranioti, S., Agathokleous, A., and Fragiadakis, M. (2013). Entropy-Based Sensor Placement Optimization for Waterloss Detection in Water Distribution Networks. Water Resour Manage. 27:4443–4468.

Colombo, A.F., Lee, P., and Karney. B. K. (2009). A Selective Literature Review of Transient-Based Leak Detection Methods. J Hydro-Environ. Res. 2(4):212-27.

Covas, D. and Ramos, H. (1999). Leakage Detection in Single Pipelines Using Pressure Wave Behaviour in Savic, D. and Walters, G. (eds) Water Industry System: Modelling and Optimisation Application. Baldock, Hertfordshire, UK.

Covas, D., Ramos, H., & Almeida, A. B. (2005). Impulse response method for solving hydraulic transients in viscoelastic pipes. XXXI IAHR Congress. Seoul, Korea.

Creaco, E., Blokker, M. and Steven Buchberger. (2017). Models for Generating Household Water Demand Pulses: Literature Review and Comparison. J. Water Resour. Plann. Manage. 143(6): 04017013.

Cressie, N. (1993). Statistics for Spatial Data. New York:Wiley.

Cugueró-Escofet, M. A., Puig, V., Quevedo, J. and Blesa, J. (2015). Optimal Pressure Sensor Placement for Leak Localisation Using a Relaxed Isolation Index: Application to the Barcelona Water Network. IFAC Papers OnLine 48(21):1108-13.

Datta, S. and Sarkar, S. (2016). A Review on Different Pipeline Fault Detection Methods. J Loss Prevent. Proc. 41:97-106.

Deagle, G., Green, A., Scrivener, J. and Tansley, N. (2007). Advanced tools for burst location. In Proceedings of the Combined International Conference of Computing and Control for the Water Industry, CCWI2007 and Sustainable Urban Water Management, SUWM2007, Leicester, UK.

Deb, K., Mohan, M. and Mishra, S. (2005). Evaluating the ε-domination based multi-objective evolutionary algorithm for a quick computation of Pareto-optimal solutions. Evolutionary Computation. 13(4):501-525.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation. 6(2):182-97.

De Joode, A. S., and Hoffman, A. (2011). Pipeline Leak Detection and Theft Detection Using Rarefaction Waves. In Proceedings of the 6th Pipeline Technology Conference, Hannover, Germany.

de Mesnard, L. (2013). Pollution models and inverse distance weighting: Some critical remarks. Computers and Geosciences. 52:459-469.

Demirci, S., Yigit, E., Eskidemir, I. H. and Ozdemir, C. (2012). Ground Penetrating Radar Imaging of Water Leaks from Buried Pipes Based on Back-Projection Method. NDT & E Int. 47: 35–42.

De Silva, D., Mashford, J. and Burn, S. (2011). Computer Aided Leak Location and Sizing In Pipe Networks. Urban Water Security Research Alliance Technical Publication No. 17 (2nd Edition).

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. Numerische Mathematik. 1(1):269-271.

DNV.GL. (2016). Synergi Water 4.9.2: User Guide. DNV GL AS. Mechanicsburg, Pennsylvania.

Fahmy, M. and Moselhi, O. (2009). Automated Detection and Location of Leaks in Water Mains Using Infrared Photography. J Perform. Constr. Facil. 24(3):242-48.

Farley, M. and Trow, S. (2003). Losses in Water Distribution Networks: A practitioner's guide to Assessment, monitoring and control. London, UK. IWA Publishing.

Farley, B., Boxall, J. B., & Mounce, S. R. (2008). Optimal Locations of Pressure Meters for Burst Detection. In Proceeding of the International Conference on Water Distribution Systems Analysis Conference, Kruger National Park, South Africa.

Farley, B., Mounce, S. R. and Boxall, J. B. (2010). Field Testing of an Optimal Sensor Placement Methodology for Event Detection in an Urban Water Distribution Network. Urban Water J. 7(6):345–56.

Farley, B., Mounce, S. R. and Boxall, J. B. (2013). Development and Field Validation of a Burst Localization Methodology. J Water Res. Plan. Man. 139(6):604-13.

Ferrante, M. and Brunone, B. (2003a). Pipe System Diagnosis and Leak Detection by Unsteady-State Tests. 1. Harmonic Analysis. Adv. Water Resources. 26(1):95–105.

Ferrante, M. and Brunone, B. (2003b). Pipe System Diagnosis and Leak Detection by Unsteady-State Tests. 2. Wavelet Analysis. Adv. Water Resources. 26(1):107-16.

Ferrante, M., Capponi, C., Collins, R., Edwards, J., Brunone, B. and Meniconi, S. (2016). Numerical Transient Analysis of Random Leakage in Time and Frequency Domains. Civ. Eng. Environ. Syst. 33(1):70–84.

Fletcher, R. and Chandrasekaran, M. (2008). SmartBall: A New Approach in Pipeline Leak Detection. In Proceedings of the 7th International Pipeline Conference, Calgary, Alberta, Canada.

Forconi, E., Kapelan, Z., Ferrante, M., Mahmoud, H., & Capponi, C. (2017). Risk-based sensor placement methods for burst/leak detection in water distribution systems. Water Sci. Technol. Water Supply. 17(6):1663-72.

Fort, T. (1948). Finite Differences and Difference Equations in the Real Domain. Oxford:Clarendon Press.

Fuchs, H.V. and Riehle, R. (1991). 10 years of experience with leakage detection by correlation analysis. Appl. Acoust. 33:1-19.

Ge, C., Guizeng W., and Hao, Y. (2008). Analysis of the Smallest Detectable Leakage Flow Rate of Negative Pressure Wave-Based Leak Detection Systems for Liquid Pipelines. Comput. Chem. Eng. 32(8):1669-80.

Giustolisi, O., and Savic, D. A. (2006). A Symbolic Data-Driven Technique Based on Evolutionary Polynomial Regression. J Hydroinform. 8(3):207-22.

Gong, J., Zecchin, A. C., Simpson, A. R. and Lambert, M. F. (2013). Frequency Response Diagram for Pipeline Leak Detection: Comparing the Odd and Even Harmonics. J Water Res. Plan. Man. 140(1):65–74.

Gong, J., Lambert, M. F., Zecchin, A. C. and Simpson, A. R. (2016). Experimental Verification of Pipeline Frequency Response Extraction and Leak Detection Using the Inverse Repeat Signal. J Hydraul. Res. 54(2):210-19.

Greenberg, J. A., Rueda, C., Hestir, E. L., Santos, M. J. and Ustin, S. L. (2011). Least Cost Distance Analysis for Spatial Interpolation. Comput. Geosci. 37(2): 272–276.

Grigg, N. S. (2006). Condition assessment of water distribution pipes. J. Infrastruct. Syst. 12(3):147-53.

Grunwell, D. and Ratcliffe, B. (1981). Location of underground leaks using the leak noise correlator. Technical Report TR 157. WRC.

Hadka, D. and Reed, P. (2013). Borg: An auto-adaptive many-objective evolutionary computing framework. Evolutionary Computation, 21(2): 231-259.

Hagos, M., Jung, D., & Lansey, K. E. (2016). Optimal meter placement for pipe burst detection in water distribution systems. J Hydroinform. 18(4):741-56.

Hallam, N. B., West, J. R., Forster, C. F., Powell, J. C., & Spencer, I. (2002). The decay of chlorine associated with the pipe wall in water distribution systems. Water Res. 36(14):3479-88.

Hart, W. E., & Murray, R. (2010). Review of Sensor Placement Strategies for Contamination Warning Systems in Drinking Water Distribution Systems. J Water Res. Plan. Man. 136(6):611–19.

Hou, Q., Ren, L., Jiao, W., Zou, P., and Song, G. (2013). An Improved Negative Pressure Wave Method for Natural Gas Pipeline Leak Location Using FBG Based Strain Sensor and Wavelet Transform. Mathematical Prob. Eng. 2013:1-8.

Hu, J., Zhang, L. and Liang, W. (2011). Detection of Small Leakage from Long Transportation Pipeline with Complex Noise. J. Loss Prevent. Proc. 24(4):449–57.

Huang, H., Tao, T. and Xin, K. (2012). Optimal pressure meters placement for bursts detection based on SOM. In Proceedings of the 14th International Conference on Water Distribution Systems Analysis, Adelaide, Australia

Hunaidi, O., Chu, W., Wang, A. and Guan, W., (2000). Detecting leaks in plastic pipes. J. Am. Water Works Ass. 92(2):82-94.

Hunaidi, O., and Wang, A., Bracken, M., Gambino, T. and Fricke, C. (2004). Acoustic Methods for Locating Leaks in Municipal Water Pipe Networks. In Proceedings of the International Conference on Water Demand Management, Dead Sea, Jordan.

Hunaidi, O., and Wang, A. (2006). A New System for Locating Leaks in Urban Water Distribution Pipes. Manage. Environ. Qual. 17(4):450–66.

International Water Association (IWA). (2000). Manual of best practice: performance indicators for water supply services. International Water Association, London, UK.

Jönsson, L. (2001). Interaction of a Hydraulic Transient with a Leak in a Pipe Flow. In Proceedings of the 14th Australasian Fluid Mechanics Conference, Adelaide, Australia.

Jung, D. and K. Lansey. (2014). Burst Detection in Water Distribution System Using the Extended Kalman Filter. In Proceedings of the 12th Conference on Computing and Control for the Water Industry, Perugia, Italy.

Jung, D. and Kevin Lansey. (2015). Water Distribution System Burst Detection Using a Nonlinear Kalman Filter. J. Water Res. Plan. Man. 141(5): 04014070.

Kapelan, Z. S., Savic, D. A., & Walters, G. A. (2005). Optimal Sampling Design Methodologies for Water Distribution Model Calibration. J. Hydraul. Eng. 131(3): 190–200.

Kapelan, Z. S., Savic, D. A., & Walters, G. A. (2003). Multiobjective Sampling Design for Water Distribution Model Calibration. J. Water Res. Plan. Man. 129(6):466–479.

Kapelan, Z.S., Savic, D.A. and Walters, G.A. (2000). Inverse Transient Analysis in Pipe Networks for Leakage Detection and Roughness Calibration. In Savic, D. A. and Walters, G. A. (eds.), Water Network Modelling for Optimal Design and Management. 1:143-159, Exeter, UK,

Kapelan, Z., Savic, D. A. and Walters, G. A. (2001). Use of Prior Information on Parameters in Inverse Transient Analysis for Leak Detection and Roughness Calibration. In Proceedings of Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges.

Kapelan, Z., Savic, D. A. and Walters, G. A. (2004). Incorporation of Prior Information on Parameters in Inverse Transient Analysis for Leak Detection and Roughness Calibration. Urban Water J. 1(2):129–43.

Kennedy, J. and Eberhart, R. (1995). Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, Western Australia.

Khan, A., Widdop, P. D., Day, A. J., Wood, A. S., Mounce, S. R. and Machell J. Artificial Neural Network Model for a Low Cost Failure Sensor: Performance Assessment in Pipeline Distribution. In Proceedings of World Academy of Science, Engineering and Technology Volume, 2006.

Kim, Y., Lee, S.J., Park, T., Lee, G., Suh, J.C. and Lee, J.M. (2016). Robust Leak Detection and Its Localization Using Interval Estimation for Water Distribution Network. Comput. Chem. Eng. 92:1-17.

Kingdom, B., Liemberger, R., Philippe, M. (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; no. 8. Washington, DC: World Bank. [Online]. Accessed 26th May 2017. Available from http://documents.worldbank.org/

Klingel, P., and Knobloch, A. (2015). A Review of Water Balance Application in Water Supply. J. Am. Water Works Ass. 107:339-50.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biol. Cybern. 43(1):59–69.

Lai, W. W .L., Chang, R. K.W., Sham, J. F.C. and Pang, K. (2016). Perturbation Mapping of Water Leak in Buried Water Pipes via Laboratory Validation Experiments with High-Frequency Ground Penetrating Radar (GPR). Tunn. Undergr. Space Tech. 52:157–67.

Laucelli, D., Romano, M., Savic, D. A. and Giustolisi, O. (2016). Detecting Anomalies in Water Distribution Networks Using EPR Modelling Paradigm. J. Hydroinform. 18(3):409–27.

LeChevallier, M. W., Gullick, R. W. and Karim, M. (2003). The Potential for Health Risks from Intrusion of Contaminants into the Distribution System from Pressure Transients. Journal of Water and Health 1, no. 1: 3–14.

Lee, P.J., Vitkovsky, J.P., Lambert, M.F., Simpson, A.R. and Liggett, J.A. (2002). Leak Detection in Pipelines Using an Inverse Resonance Method. In Proceedings of the Conference on Water Resources Planning & Management, Roanoke, Virginia, USA

Lee, P.J., Vitkovsky, J.P., Lambert, M.F., Simpson, A.R. and Liggett, J.A. (2005a). Frequency domain analysis for detecting pipeline leaks. J. Hydraul. Eng. 131(7):596-604.

Lee, P.J., Vitkovsky, J.P., Lambert, M.F., Simpson, A.R. and Liggett, J.A (2005b). Leak Location Using the Pattern of the Frequency Response Diagram in Pipelines: A Numerical Study. J. Sound Vib. 284(3–5):1051-73.

Lee, P.J., Vitkovsky, J.P., Lambert, M.F., Simpson, A.R. and Liggett, J.A. (2007). Leak Location in Pipelines Using the Impulse Response Function. J. Hydraul. Res. 45(5):643-52.

Liemberger, R., and Farley, M. (2004). Developing a non-revenue water reduction strategy: investigating and assessing water losses. In Proceedings of the International Water Association 4th World Water Congress, Marrakech, Morocco.

Licciardi, S.V. (1998). Quantification and location of leaks by internal inspection. Tunn. Undergr. Sp. Tech., 13(2):5-15.

Liggett, J.A. and Chen, L.C. (1994). Inverse transient analysis in pipe networks. J. Hydraul. Eng. 120(8):934-55.

Li, R., Huang, H., Xin, K. and Tao, T. (2015). A Review of Methods for Burst/Leakage Detection and Location in Water Distribution Systems. Water Science & Technology: Water Supply 15(3):429-41.

Loth, J., Morris, G. J., Palmer, G. M., Guiler, R. and Mehra. M. (2003). Technology Assessment of on-Line Acoustic Monitoring for Leaks/Infringements in Underground Natural Gas Transmission Lines. West Virginia University, USA.

Mamdani, E. H. and Assilian, S. (1975). An Experiment in Linguistic Synthesis with a Fuzzy Logic Controller. Int. J. Man Mach. Stud. 7(1): 1–13.

Mashford, J., De Silva, D., Marney, D. and Burn, S. (2009). An Approach to Leak Detection in Pipe Networks Using Analysis of Monitored Pressure Values by

Support Vector Machine. In Proceedings of the 3rd International Conference on Network and System Security, Gold Coast, Australia.

Machell, J., Mounce, S. R., and Boxall, J. B. (2010). Online modelling of water distribution systems: A UK case study. Drink. Water Eng. and Sci. 3(1): 21-27.

Metje, N., Atkins, P. R., Brennan, M. J., Chapman, D. N., Lim, H. M., Machell, J., Muggleton, J. M., Pennock, S., Ratcliffe, J., Redfern, M., Rogers, C. D. F., Saul, A. J., Shan, Q., Swingler, S. and Thomas, A. M. (2007). Mapping the Underworld – State-of-the-Art Review. Tunn. Undergr. Sp. Tech. 22(5–6):568–86.

Misiunas, D., Lambert, M. F., Simpson, A. R. and Olsson, G. (2005). Burst Detection and Location in Water Transmission Pipelines. In Proceedings of the 7th Conference on Water Distribution Systems Analysis, Anchorage, Alaska.

Morrison, J., Tooms, S. and Rogers, D. (2007). District Metered Areas Guidance Notes. [Online]. [Accessed 21st May 2017]. Available from: http://operationandmaintenance.net/uploads/IWA%20Toolboxes/OM%20Toolbox/Images/EOM%20tool2%20DMA%20Guidance%20Notes.pdf

Mounce, S. R., and Machell, J. (2006). Burst Detection Using Hydraulic Data from Water Distribution Systems with Artificial Neural Networks. Urban Water J. 3(1):21–31.

Mounce, S. R., Day, A. J., Wood, A. S., Khan, A., Widdop, P. D. and Machell, J. (2002). A Neural Network Approach to Burst Detection. Water Sci. Technol. 45(4–5):237-46.

Mounce, S. R., Khan, A., Wood, A. S., Day, A. J., Widdop, P. D. and Machell, J. (2003). Sensor-Fusion of Hydraulic Data for Burst Detection and Location in a Treated Water Distribution System. Inform. Fusion. 4(3): 217–29.

Mounce, S. R., Boxall, J. B. and Machell, J. (2007). An Artificial Neural Network/Fuzzy Logic system for DMA flow meter data analysis providing burst identification and size estimation. In Proceedings of the Combined International Conference of Computing and Control for the Water Industry, CCWI2007 and Sustainable Urban Water Management, SUWM2007, Leicester, UK.

Mounce, S. R., Boxall, J. B. and Machell, J. (2010). Development and Verification of an Online Artificial Intelligence System for Detection of Bursts and Other Abnormal Flows. J. Water Res. Plan. Man. 136(3):309-18.

Mounce, S. R., Mounce, R. B. and Boxall, J. B. (2011). Novelty Detection for Time Series Data Analysis in Water Distribution Systems Using Support Vector Machines. J. Hydroinform. 13(4):672-86.

Mpesha, W., Gassman, S. L. and Chaudhry, M. H., 2001. Leak detection in pipes by frequency response method. J. Hydraul. Eng. 127(2):134-46.

Mukai, H., & Polak, E. (1976). On the implementation of reduced gradient methods. In J. Cea (Ed.), Optimization Techniques Modeling and Optimization in the Service of Man Part 2. 41:426–437. Springer, Berlin, Heidelberg.

Mukherjee, J., and Narasimhan, S. (1996). Leak Detection in Networks of Pipelines by the Generalized Likelihood Ratio Method. Ind. Eng. Chem. Res. 35(6):1886-93.

Mukkamala, S., Janovski, G. and Sung, A.H., 2002. Intrusion Detection Using Neural Networks and Support Vector Machines. Proceedings of the 2002 International Joint Conference on Neural Networks, Honolulu, USA.

Mutikanga, H. E., Sharma, S. K. and Vairavamoorthy, K. (2013). Methods and Tools for Managing Losses in Water Distribution Systems. J. Water Res. Plan. Man. 139(2): 166-74.

National Resources Defence Council (NRDC). (2010). Water Facts. [Online]. [Accessed 21st May 2017]. Available from: http://www.nrdc.org/global-Warming/watersustainability/.

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A., Zechman, E. State of the Art for Genetic Algorithms and Beyond in Water Resources Planning and Management. J. Water Res. Plan. Man. 136(4):412-32.

Nixon, W. and Ghidaoui, M.S. (2007). Numerical sensitivity study of unsteady friction in simple systems with external flows. J. Hydraul. Eng. 133(7):736-49.

Oakland, J. S. (2003). Statistical Process Control. 5th ed. Oxford. Butterworth-Heinemann.

Ofwat. (2017). The guaranteed standards scheme (GSS): summary of standards and conditions. Available at: https://www.ofwat.gov.uk/wp-content/uploads/2017/03/The-guaranteed-standards-scheme-GSS-summary-of-standards-and-conditions.pdf (Accessed 09th June 2020).

Okeya, I., Kapelan, Z., Hutton, C. and Naga, D. (2014). Online Burst Detection in a Water Distribution System Using the Kalman Filter and Hydraulic Modelling. Proceedings of the 16th International Conference on Water Distribution Systems Analysis, Bari, Italy.

Ovarro. (2020). Datective. [ONLINE]. [Accessed 15th December 2020]. Available from https://ovarro.com/en/europe/solutions/analytics/on-premise/datective/.

Paez, D., Suribabu, C.R. and Filion, Y. (2018). Method for Extended Period Simulation of Water Distribution Networks with Pressure Driven Demands. Water Resources Management. 32(8):2837-2846.

Palau, C. V., Arregui, F. and Ferrer, A. (2004). Using Multivariate Principal Component Analysis of Injected Water Flows to Detect Anomalous Behaviors in a Water Supply System-a Case Study. Water Sci. and Technol.: Water Supply 4(3):169-82.

Palau, C.V., Arregui, F. J. and Carlos, M. (2012). Burst Detection in Water Networks Using Principal Component Analysis. J. Water Res. Plan. Man. 138:47-54.

Pearson, D. and Trow, S. (2005). Calculating Economic Levels of Leakage. Proceedings of the IWA Water Loss Taskforce Leakage 2005 Conference, Halifax, Nova Scotia.

Pérez, R., Puig, V., Pascual, J., Peralta, A., Landeros, E., & Jordanas, Ll (2009). Pressure sensor distribution for leak detection in Barcelona water distribution network. Water Supply, 9(6):715-21.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E., & Peralta, A. (2011). Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. Control Eng. Pract. 19(10):1157-67.

Phair, D.A. (1997). Errors occur often in municipal flowmetering. INTECH, 44(9). 54-57.

Pilcher, R., 2003. Leak detection practices and techniques: a practical approach. Water 21:44-45.

Pilcher, R., Hamilton, S., Chapman, H., Field, D., Ristovski, B. and Strapely, S. (2007). Leak Location and Repair Guidance Notes. IWA Water Loss Task Force. https://www.pseau.org/outils/ouvrages/iwa_guidance_notes_on_leak_detection_and_repair_2007.pdf

Pudar, R. S. and Liggett, J. A. (1992). Leaks in pipe networks. J. of Hydraul. Eng. 118(7):1031-46.

Pudil, P., Novovičová, J., & Kittler, J. (1994). Floating search methods in feature selection. Pattern Recognit. Lett. 15(11):1119-25.

Puleo V, Freni G, La Loggia G (2018) Pressure sensors positioning for leakages detection under uncertain demands. Proceedings of the 13th international conference on Hydroinformatics, Palermo, Italy

Puust, R., Kapelan, Z., Savic, D. A. and Koppel, T. (2006). Probabilistic Leak Detection in Pipe Networks Using the SCEM-UA Algorithm. In Proceedings of the Symposium on Water Distribution Systems Analysis Symposium, Cincinnati, USA.

Puust, R., Kapelan, Z., Savic, D. A. and Koppel, T. (2010). A Review of Methods for Leakage Management in Pipe Networks, Urban Water J. 7(1):25-45.

Rashid, S., Qaisar, S, Saeed, H., and Felemban, E. (2014). A Method for Distributed Pipeline Burst and Leakage Detection in Wireless Sensor Networks Using Transform Analysis. Int. J. Distrib. Sens. N. 10(7):939657.

Raei, E., Nikoo, M. R., Pourshahabi, S., & Sadegh, M. (2018). Optimal joint deployment of flow and pressure sensors for leak identification in water distribution networks. Urban Water J., 15(9):837-46.

Raei, E., Shafiee, M. E., Nikoo, M. R., & Berglund, E. (2019). Placing an ensemble of pressure sensors for leak detection in water distribution networks under measurement uncertainty. J. Hydroinform. 21(2):223-39.

Robinson, T. P. and Metternicht, G. (2006). Testing the performance of spatial interpolation techniques for mapping soil properties. Computers and Electronics in Agriculture. 50(2):97-108.

Romano, M., Kapelan, Z., and Savić, D. A. (2010). Real-time leak detection in water distribution systems. Proceedings of the Conference on Water Distribution System Analysis, Tucson, Arizona, USA

Romano, M., Kapelan, Z. and Savic, D. A. (2011). Burst detection and location in water distribution systems. Proceedings of the World Environmental and Water Resources Congress 2011: Bearing Knowledge for Sustainability, Palm Springs, California, USA.

Romano, M., Kapelan, Z. and Savic, D. A. (2012). Near Real-Time Detection and Approximate Location of Pipe Bursts and Other Events in Water Distribution Systems. PhD Thesis, University of Exeter, Exeter, UK.

Romano, M., Kapelan, Z. and Savic, D. A. (2013). Geostatistical Techniques for Approximate Location of Pipe Burst Events in Water Distribution Systems, J. Hydroinform. 15(3):634–651.

Romano, M., Kapelan, Z. and Savic, D. A. (2014a). Automated Detection of Pipe Bursts and Other Events in Water Distribution Systems, J. Water Res. Plan. Man. 140(4):457-67.

Romano, M., Kapelan, Z. and Savic, D. A. (2014b). Evolutionary algorithm and expectation maximization strategies for improved detection of pipe bursts and other events in water distribution systems. J. Water Res. Plan. Manage. 140(5):572-84.

Romano, M., Woodward, K. and Kapelan, Z. (2016). Statistical Process Control Based System for Approximate Location of Pipe Bursts and Leaks in Water Distribution Systems. In Proceedings of the 18th Conference on Water Distribution Systems Analysis, Cartagena, Colombia.

Romano, M. (2019). 'Review of Techniques for Optimal Placement of Pressure and Flow Sensors for Leak/Burst Detection and Localisation in Water Distribution Systems', in Dawei H., Mounce, S., Scozzari, A., Soldovieri, F., and Solomatine, D. (eds.), ICT for Smart Water Systems: Measurements and Data Science, Hdb Env Chem.

Sarrate, R., Blesa, J., Nejjari, F. and Quevedo, J. (2014). Sensor Placement for Leak Detection and Location in Water Distribution Networks. Water Science & Technology: Water Supply 14(5):795-803.

Savic, D. A., Bicik, J., Morley, M. S., Duncan, A., Kapelan, Z., Djordjević, S., & Keedwell, E. C. (2013). Intelligent urban water infrastructure management. Journal of the Indian Institute of Science. 93(2): 319-336.

Savic, D. A., Kapelan, Z. S., & Jonkergouw, P. M. R. (2009). Quo vadis water distribution model calibration? Urban Water J. 6(1):3–22.

Savic, D. A. and Walters, G. A. (1997). Genetic algorithms for least-cost design of water distribution networks. Journal of Water Resources Planning and Management. 123(2):67-77.

Sheltami, T. R., Bala, A. and Shakshuki., E. M. (2016). Wireless Sensor Networks for Leak Detection in Pipelines: A Survey. J. Amb. Intel. Hum. Comp. 7(3):347-56.

Shuqing, Z., Tianye G., Hong X., Guangpu H., and Zhongdong W., (2009). Study on New Methods of Improving the Accuracy of Leak Detection and Location of Natural Gas Pipeline. In Proceedings of the International Conference on Measuring Technology and Mechatronics Automation, Zhangjiajie, China.

Silva, R. A., Buiatti, C. M., Cruz, S. L. and Pereira, J. A. F. R. (1996). Pressure wave behaviour and leak detection in pipelines. Comput. Chem. Eng. 20(1):491-96.

Skipworth, P. J., Saul, A. J. and Machell, J. (1999). Predicting water quality in distribution systems using artificial neural networks. Proceedings of the Institution of Civil Engineers: Water, Maritime and Energy. 136(1):1-8.

Soldevila, A., Fernandez-Canti, R. M., Blesa, J., Tornil-Sin, S. and Puig, V. (2017). Leak localization in water distribution networks using Bayesian classifiers. Journal of Process Control. 55:1-9.

Soldevila, A., Blesa, J., Fernandez-Canti, R. M., Tornil-Sin, S. and Puig, V. (2019). Data-Driven Approach for Leak Localization in Water Distribution Networks Using Pressure Sensors and Spatial Interpolation. Water, 11(7):1500.

Sophocleous, S., Savić, D., & Kapelan, Z. (2019a). Leak Localization in a Real Water Distribution Network Based on Search-Space Reduction. J. Water Res. Plan. Man. 145(7):04019024.

Sophocleous, S., (2019b). Development of the Next Generation of Water Distribution Network Modelling Tools Using Inverse Methods. EngD Thesis, University of Exeter, Exeter, UK.

SPA Future Thinking. (2013). CCWater Leakage Study: Research into customer perceptions of leakage. [Online]. [Accessed 21st May 2017]. Available from: https://www.ccwater.org.uk/wp-content/uploads/2013/12/Research-into-customer-perceptions-of-leakage.pdf

Srirangarajan, S., Iqbal, M., Lim, H.B., Allen, M., Preis, A., and Whittle, A. (2010a). Water Main Burst Event Detection and Localization. In Proceeding of the Conference on Water Distribution System Analysis, Tucson, Arizona, USA

Srirangarajan, S., Iqbal, M., Allen, M., Preis, A., Fu, C., Girod, L., Wong, K. J., Lim, H. B. and Whittle, A. (2010b). Real-time burst event detection in water distribution systems. In Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, Stockholm, Sweden.

Stampolidis, A., Soupios, P., Vallianatos, F. and Tsokas, G. N. (2003). Detection of leaks in buried plastic water distribution pipes in urban places - A case study, Proceedings of the 2nd International Workshop on Advanced Ground Penetrating Radar, Delft, Netherlands.

Steffelbauer, D., Neumayer, M., Günther, M. and Fuchs-Hanusch, D. (2014). Sensor Placement and Leakage Localization Considering Demand Uncertainties. In Proceedings of the 16th Conference on Water Distribution System Analysis, Bari, Italy.

Steffelbauer, D., and Fuchs-Hanusch, D. (2016). Efficient Sensor Placement or Leak Localization Considering Uncertainties. Water Resour. Manage. 30(14):5517-33.

Stoianov, I., Karney, B., Covas, D., Maksimović, C. and Graham, N. (2001). Wavelet Processing of Transient Signals for Pipeline Leak Location and Quantification. Proceedings of the International Conference on Computing and Control for the Water Industry, Leicester, UK.

Storn, R. and Price, K. (1997). Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. J. Global Optim. 11(4):341-59.

Stroud, K. A. and Booth, D. J. (2011). Advanced Engineering Mathematics. 5th Ed. New York:Palgrave Macmillan.

Tabesh, M., Asadiyani Yekta, A. H. and Burrows, R. (2009). An Integrated Model to Evaluate Losses in Water Distribution Systems. Water Resour. Manage. 23(3): 477–92.

Tang, K.W., Karney, B. and Brunone, B. (2001). Leak detection using inverse transient calibration and GA - some early success and future challenges. In: Water Software Systems: theory and applications by (Eds) Ulanicki, B., Coulbeck, B. and Rance, J. 2:55-64.

Tian, C. H., Yan, J. C., Huang, J., Wang, Y., Kim, D. S. and Yi, T. (2012). Negative pressure wave based pipeline Leak Detection: Challenges and algorithms, Proceedings of 2012 IEEE International Conference on Service Operations and Logistics, and Informatics, Suzhou, China.

Uber, J. and Hatchett, S. (2013). OpenWaterAnalystics: EPANET-Matlab. [ONLINE]. [Accessed 15th December 2020]. Available from https://github.com/OpenWaterAnalytics/epanet-matlab.

UKWIR. (2013). Cost Benefit Analysis of Ubiquitous Data Collection in Water Distribution - CBAScenarios. [ONLINE]. [Accessed 15th December 2020] Available from https://ukwir.org/reports/13-DW-12-2/66678/Cost-Benefit-Analysis-of-Ubiquitous-Data-Collection-in-Water-Distribution--CBA-Scenarios.

Van Thienen, P. (2013). A Method for Quantitative Discrimination in Flow Pattern Evolution of Water Distribution Supply Areas with Interpretation in Terms of Demand and Leakage. J. Hydroinform. 15(1):86-102.

Van Thienen, P. and Montiel, F. (2014). Flow Analysis and Leak Detection with the CFPD Method in the Paris Drinking Water Distribution System. In Proceedings of the 11th International Conference on Hydroinformatics, New York, USA.

Vitens, Acciona, Thames Water and Lille University. (2018). Demonstration of integrated smart water supply solutions at 4 sites across Europe. [ONLINE]. [Accessed 09th April 2021] Available from https://cordis.europa.eu/project/id/619024/reporting.

Vitkovsky, J. P., Simpson, A. R. and Lambert, M. F. (2000). Leak detection and calibration using transients and genetic algorithms. J. Water Res. Plan. Man. 126(4):262-65.

Vitkovsky, J. P., Liggett, J. A., Simpson, A. R. and Lambert, M. F. (2003). Optimal measurement site locations for inverse transient analysis in pipe networks. J. Water Res. Plan. Man. 129(6):480-92.

Von Luxburg, U. (2007). A tutorial on spectral clustering. Stat. Comput. 17(4), 395-416.

Wang, X. J., Lambert, M. F., Simpson, A. R., Vitkovsky, J. P. (2001). Leak Detection in Pipelines and Pipe Networks: A Review. In Proceedings of the 6th Conference on Hydraulics in Civil Engineering, Hobart, Australia.

Wang, X. J., Lambert, M. F., Simpson, A. R., Liggett, J. A. and Vitkovsky, J. P. (2002). Leak detection in pipelines using the damping of fluid transients. J. Hydraul. Eng. 128(7):697-711.

Wang, Q., Savic, D. A. and Kapelan, Z. (2017). GALAXY: A new hybrid MOEA for the optimal design of Water Distribution systems, Water Resour. Res. 53:1997-2015.

Wang, P., Zecchin, A. C., Maier, H. R., Zheng, F. and Newman, J. P. (2020). Do Existing Multiobjective Evolutionary Algorithms Use a Sufficient Number of Operators? An Empirical Investigation for Water Distribution Design Problems, Water Resour. Res. 56 (5): e2019WR026031

Wu, Y. and Liu, S. (2017). A review of data-driven approaches for burst detection in water distribution systems. Urban Water J. 14(9):972-83.

Wu, Z. Y. and Sage, P. (2006). Water Loss Detection via Genetic Algorithm Optimization-based Model Calibration. In Proceedings of the International Symposium on Water Distribution System Analysis, Cincinnati, USA.

Wu, Z. Y. and Sage, P. (2007). Pressure dependent demand optimization for leakage detection in water distribution systems. In Proceedings of the Combined International Conference of Computing and Control for the Water Industry, CCWI 2007 and Sustainable Urban Water Management, SUWM2007, Leicester, UK.

Wu, Z. Y., Sage, P. and Turtle, D. (2010). Pressure-dependent leak detection model and its application to a district water system. J. Water Res. Plan. Man., 136(1):116-28.

Wu, Z. Y., Wang, Q, Butala, S. and Mi T. (2011). Darwin Optimization Framework User Manual. Bentley Systems Incorporated, Watertown, USA.

Wu, Z. Y. and Song, Y. (2012). Optimising pressure logger placement for leakage detection and model calibration. Proceedings of the 14th International Conference on Water Distribution Systems Analysis, Adelaide, Australia

Ye, G. and Fenner, R. A. (2011). Kalman filtering of hydraulic measurements for burst detection in water distribution systems. J. Pipeline Syst. Eng. and Practice. 2(1):14-22.

Zimmerman, D., Pelvic, C., Ruggles, A. and Armstrong, M. P. (1999). An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting, Math. Geol. 31(4),:375–390.