Deriving further information from the leak signal in water distribution pipes

Joseph David Butterfield

August 2018

A THESIS SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF ENGINEERING DOCTORATE

The University of Sheffield

Dedication

To my parents, thank you for all of your support and for helping me find my own way. To Marika, I became a better person the day I met you. Thank you for believing in me.

Acknowledgements

A number of people have supported me in the development of this thesis and they all deserve my gratitude. Firstly, I would like to thank my Supervisors Professor Stephen Beck and Dr Richard Collins for their guidance and support throughout. In the Water Laboratory, I would like to thank all the staff, in particular Paul Osborne and Alex Cargill whom assisted me with building the pipe rig. Also at the University of Sheffield I would like to thank Dr Anton Krynkin for meetings and help with some of the technical content of my work. My office colleagues, Dr Stewart Husband, Dr Will Shepard and Ms Sally Jones also deserve my gratitude for enabling me to enjoy my time spent in the office. I was also lucky enough to be able to collaborate with some external partners. Of these, Dr Vivana Muranne at the Universidad de Chile, assisted with developing and understanding the Least Squares-Support Vector Machine model presented in Chapter 6 and deserves my full gratitude. Thanks also go to Mr Gregory Meyers at the University of Exeter, whom helped me with some of the modelling presented in Chapter 7.

Special thanks goes to my industrial sponsors, without which this research would have never been carried out. Thank you to Tim Evans (Thames Water), Paul Taylor (Severn Trent Water), Mark Haffrey (Scottish Water) who contributed during project steering group meetings. A particular thanks goes to Claire Gowdy and Jason Plunkett at Northumbrian Water who helped with facilitating my visits to field sites, administrative support and general guidance. A special thanks goes to Mr Dennis Dellow at Northumbrian Water who helped with driving the project in a direction which would provide results of high value to industry and for technical support throughout.

I would also like to thank my colleagues on the 'STREAM' programme, in particular Cohort V. Attending such a programme provided me with a vast network of support.

Abstract

Leaking pipes from water distribution systems are a huge issue which has maintained worldwide attention. A common method of detecting leaks is through the cross correlation of leak vibro-acoustic emission signals, created as the water discharges through the hole/crack and recorded by acclerometers or hydrophones placed either side of a suspected leak location. It is thought that a number of factors influence leak signals, however there has been little comprehensive study in controlled conditions evaluating how and what way these factors influence the characteristics of leaks signals. Moreover, during the process of leak noise correlation, accelerometers and hydrophones are recording information about the leak contained in the leak signal that is currently not understood. Knowledge of these factors would be highly beneficial to water companies in order to allow for prioritisation of leak repair.

The research presented herein aimed to derive a method in order to predict the flow rate, area and shape of a leak using the vibo-acoustic emission signal recorded when performing cross correlation. A unique methodology was developed which allowed for the isolation of individual physical variables and how this can influence a leak signal. Specifically, the study focused on developing a fundamental understanding of how leak flow rate, area, shape, pipe material and backfill type influenced the characterstics of leak signals.

The results showed that the leak flow rate, shape, backfill types and pipe material all influence the leak signal. The influence of leak area on the leak signal appeared negligible when leak flow rates were standardised. Signal processing and machine learning algorithms were applied to the leak signals and the results showed that it was possible to predict leak flow rate regardless of leak area, leak shape and backfill type. Moreover, alternate algorithms showed that it was possible to predict leak shape and leak area from the vibro-acoustic emission signal. This research has therefore presented a useful and valid tool to predict leak flow rate, leak area and leak shape which allows water companies to prioritise leak repair and maintenance activities, providing an opportunity to reduce the volume of water lost through leaks by repairing the larger flow rate leaks first. Whilst this method shows effective results, it does not provide an exhaustive comparison of the number of algorithms and techniques which may also make similar predictions.

List of publications

The research incorporated within this thesis has already contributed a number of contributions to knowledge through published journal articles, international conferences and presentation to the wider water industry. Details of these are below. Parts of this thesis are also included in the published papers.

Published journal papers

- Butterfield, J.D., Meruane, V., Collins, R.P., Meyers, G. and Beck, S.B.M. (2017) Prediction of leak flow rate in plastic water distribution pipes using vibro-acoustic measurements, *Structural Health Monitoring*, 359, pp. 40-55.
- Butterfield, J.D., Krynkin, A., Collins, R.P. and Beck, S.B.M. (2017) Experimental investigation into vibro-acoustic emission signal processing techniques to quantify leak flow rate in plastic water distribution pipes, *Applied Acoustics*, 119, pp. 146-155.
- Butterfield, J.D., Collins, R.P. and Beck, S.B.M. (2018) A comparative study on the influence of pipe material on the transmission of leak vibro-acoustic signals in real complex water distribution systems: a case study, *ASCE Journal of Pipeline Engineering and Practice*, 9(3), pp. 1-17.
- Butterfield, J.D., Meyers, G., Meruane, V., Collins, R.P. and Beck, S.B.M. (2018) Experimental investigation into techniques to predict leak shapes in water distribution systems using vibration measurements, *Journal of Hydroinformatics*, 20(4), pp. 815828..

Published conference papers and presentations

- Butterfield, J.D., Collins, R.P. and Beck, S.B.M. (2015) Improving the leak location accuracy by optimising the time delay measurement of acoustic emission from leaks in MDPE pipe. *Proceedings of the 22$^{\mathrm{nd}}$ International Congress on Sound and Vibration*, Florence, Italy.
- Butterfield, J.D., Collins, R.P. and Beck, S.B.M. (2015) Feature extraction of leak signals in plastic water distribution pipes using the wavelet transform, *Proceedings of the ASME 2015 International Mechanical Engineering Congress and Exposition*, Houston, Texas.
- Butterfield, J.D., Collins, R.P. and Beck, S.B.M. (2017) Experimental investigation into the influence of backfill types on the vibro-acoustic characteristics of leaks in MDPE pipe, *Procedia Engineering*, Water Distribution Systems and Analysis conference, Cartegena, Colombia, pp. 311-318.

- Butterfield, J.D., Collins, R.P. and Beck, S.B.M. (2017) Influence of Pipe Material on the Transmission of Leak Vibro-acoustic Signals in Water Distribution Systems, 8^{th} *International Conference on Acoustic Emission*, Kyoto, Japan.

- Butterfield, J.D., Meruane, V., Collins, R.P., Meyers, G. and Beck, S.B.M. (2017) Application of least squares support vector machines and acoustic measurements for leak flow rate prediction, *Proceedings of the Computing and Control for the Water Industry*, Sheffield, UK.

Contents

List of Tables

List of Figures

List of abbreviations

AE	Acoustic Emission
AWWA	American Water Works Association
AC	Asbestos Cement
BCC	Basic Cross Correlation
CI	Cast Iron
C	Copper
DI	Ductile Iron
DMA	District Metered Area
EEMD	Ensemble Empirical Mode Decomposition
EMD	Empirical Mode Decomposition
FE	Fire Hydrant
GCC	Generalised Cross Correlation
GPR	Ground Penetrating Radar
HDPE	High Density Polyethylene
ICA	Independent Component Analysis
IEPE	Integrated Electronics Piezo-Electric
IMF	Intrinsic Mode Functions
KPLSR	Kernel Partial Least Squares Regression
LS-SVM	Least Squares Support Vector Machines
LR	Logistic Regression
ML	Maximum Likelihood
MDPE	Medium Density Polyethylene
mV	Millivolts
MLR	Multiple Linear Regression
NRW	Non-Revenue Water
Ofwat	Office of Water
PLSR	Partial Least Squares Regression
PHAT	Phase Transform
PE	Polyethylene
PSD	Power Spectral Density
PCA	Principal Component Analysis
RBF	Radial Basis Function
RF	Random Forest
RMS	Root Mean Square
RMSE	Root Mean Square Error

SCOT	Smoothed Coherence Transform
SELL	Sustainable Economic Level of Leakage
SNR	Signal to Noise Ratio
SVM	Support Vector Machines
WDS	Water Distribution Systems
VAE	Vibro-Acoustic Emission
V	Volts

Nomenclature

d	Total distance between sensor 1 and sensor 2 (m)
t	Time (s)
L_1	Distance from sensor 1 (m)
L_2	Distance from sensor 2 (m)
c	Wavespeed (m/s)
τ_{delay}	Time delay, lag of time (s)
x_1	Leak signal and background noise recorded at sensor 1
x_2	Leak signal and background noise recorded at sensor 2
S	Leak signal
n_1	Background noise recorded at sensor 1
n_2	Background noise recorded at sensor 2
c	Wavespeed (m/s)
τ_{delay}	Time delay, lag of time (s)
$R_{x_1x_2}$	Basic cross correlation function between two signals x_1 and x_2
E	Expectation operator
$\rho(\tau)$	Normalised cross correlation coefficient
c_f	Free fluid wavespeed
B	Bulk density
E	Young's Modulus of Elasticity
a	Pipe radius
h	Pipe wall thickness
Q	Leak flow rate
C_d	Discharge coefficient
A	Hole area
g	Gravity constant
h	Pressure (m)
ϖ	Leakage coefficient
ν	Pressure
α	Leakage exponent
δ	Attenuation factor
n	Coefficient for damping in the pipe wall
ω	Frequency
$H_{F(jw)}$	Transfer function
G	Filters gain
ϑ	Filter order

$\psi_{SCOT}(\omega)$	SCOT weighting for GCC filter
$\psi_{PHAT}(\omega)$	PHAT weighting for GCC filter
$\psi_{ML}(\omega)$	ML weighting for GCC filter
$\gamma^2{}_{x_1x_2}$	Coherence function x_1 and x_2
$\psi_g(\omega)$	Weighting functions of $\psi_{SCOT}(\omega)$, $\psi_{PHAT}(\omega)$, $\psi_{ML}(\omega)$ using GCC
F^{-1}	Fourier transform
$S_{x_1x_2}$	Cross Spectral Density
y	Leak flow rate (only linear and non-linea regression)
e	Residuals (linear and non-linear regression)
β	Slope parameter (multiple linear and non-linear regression)
$X_1...X_n$	Explanatory variables (Logistic Regression)
Y	Binary variable (Logistic Regression)
$x_i \epsilon R^n$	Data from given sample $(x_i, y_i)_i^m = 1$
J	Objective function
C	Regularisation parameter ($\gamma and \sigma$)
k	Slack variable
b	Bias function
$\frac{1}{2}w^Tw$	Flatness measurement function
φ	Non-linear function mapping input into a higher dimensional space (LS-SVM)
α_i	Lagrange Multipliers
AE_{RMS}	RMS value of x_1
C_1	Constant relating to valve material, attenuation and sensor
D	Valve size
α	Sound velocity
ρ	Fluid density
C_v	Valve flow coefficient
S	Fluid specific gravity
ZP	Pressure drop across the valve
$P[k]$	Power Spectral Density
T	Sampling period
$X[k]$	Discrete Fourier Transform

Chapter 1

Motivation

1.1 Introduction

Pipelines have been the primary method of transporting water for centuries, with some pipes dating back to the 15^{th} century (Tayefi, 2014). Nowdays, the Water Distribution System (WDS) of the United Kingdom (UK) is huge, reportedly carrying over 8500 Mega-Litres of water per day (Fox, 2016). Water loss from from these leaking pipes has been a problem since their installation. Leakage has a number of negative consequences, including economic loss due to non-revenue water (Colombo and Karney, 2002), environmental effects and water quality issues due to risk of contaminant ingress (Fox et al., 2014). Coupled with challenges such as rising populations and climate change which exert additional pressures on water resources, leaking pipes are a major issue worldwide.

Leakage levels remain a major challenge for water companies. Between 2009-2011, leakage levels in the UK were estimated to be 133.1 litres per day per property, around 24% (Ofwat, 2010) of the drinking water entering WDS. However, the volume of water loss varies around the world: in some less developed counties, leakage levels can be >50%, whereas in some more developed countries with improved infrastructure leakage levels can be <5% (AWWA, 1987; Almeida et al., 2014a). The level of leakage depends on a number of factors, such as network characteristics and age, as well as operational management aspects (such as pressure management) (Ofwat, 2010). Generally, driven by customer, regulatory and economic pressures, water companies in the UK have lowered leakage levels through pressure management and active leak detection/reduction programmes using a variety of techniques. In order to prevent increasing leakage rates, water companies have, where cost-effective, replaced water mains with Polyethylene (PE) pipes. PE pipes are assumed to leak less (Fox, 2016) and are low cost whilst also being more resistant to corrosion (Pal, 2008) compared to metalic pipes. However, PE pipes are far from leak free (Fox, 2016) and due to the challenges around leak detection on plastic pipe, driving down leakage rates further represents a major challenge. Although water companies have pro-actively attempted to drive down leakage rates further, in some cases leakage rates have plateaued and not reduced in the last four years (WWI, 2017), or have in some cases increased despite greater fears of drought (Guardian, 2017). Failure to meet leakage targets set by the UK water industry regulator Office of Water (Ofwat) can have significant implications for water companies. For example, Thames Water based in Southern

England, have recently been penalised £8.85 million (BBC, 2018) and £120 million for not achieving leakage targets during the year 2016-2017 and 2017-2018 respectively.

A variety of tools to locate leaks are at a water companies disposal, and leaks are often identified visually by customers reporting water noticed on the ground. However, the unnoticed leaks which cannot be identified visually constitute a huge level of water loss. In these cases, a leak is normally found at the District Metered Area (DMA) scale by analysing trends in customer night use. At the individual pipeline scale, a variety of methods are available in order to identify the location of the leak, with varying levels of success. The tool most often used in the UK is that of leak noise correlation. As the water discharges through a leak, a noise is created due to turbulence and/or cavitation. This noise then travels along the pipe in the fluid and also on the pipe wall. Sensors placed either side of the leak with the addition of signal processing are often able to derive the location of the leak. Whilst proven to be successful on metallic pipes, leak noise correlation methods have lower success rates on plastic pipes due to the rapid attenuation of the leak signal (Almeida et al., 2014a).

To further reduce leakage rates, the UK water regulator, Ofwat) introduced the Sustainable Economic Level of Leakage (SELL). The aim of the SELL mechanism is to balance the cost of repair with the costs of leakage. The SELL mechanism also aims to take in to account of the social and environmental factors of leakage (CIWEM, 2017). In some cases it is therefore cheaper to allow a leak to continue to discharge rather than repair the leak. Therefore, the decision whether or not to repair a leak can largely depend on the leak flow rate. Moreover, the knowledge of the leak flow rate can allow for the prioritisation of leak repair activities and company resources. The prioritisation of operational workloads will therefore reduce the costs of finding a leak and thereby reduce SELL, ultimately changing the economics of leakage. However, no tool exits whereby the leak flow rate can be accurately quantified.

There are numerous benefits associated with the knowledge of the leak flow rate, area and shape which will allow water companies to priorities leak repair and maintenance strategies. The overall aim of this research is to therefore develop a tool that quantifies the leak flow rate, leak area and shape from the leak's VAE signal.

1.2 Research aim

This research will attempt to derive further information from the VAE signal produced by the leak and its interaction with other variables, whether this be using vibration or acoustic sensors. In particular, it is desirable to derive from the leak signal a method which can characterise leak flow rate, leak area and leak shape shape using existing acoustic and vibration sensors. In order to do this, a thorough understanding of how the different factors influence the leak signal is required. The methodology in order to develop this understanding will combine unique experimental techniques and methods used in other disciplines, such as machine learning and signal processing in order to achieve this.

1.3 Research questions and objectives

The following section details a number of research questions (shown in bold text) and research objectives.

1. **What is the current state of research knowledge investigating the variables influencing a leak noise and the accuracy of leak detection?** Review the existing body of literature regarding pipeline leakage detection, characteristics of the leak noise and methods to classify/predict leak flow rate, leak area and leak shape. Also review signal processing and machine learning techniques in order to achieve a classification and prediction model of the aforementioned variables from the leak VAE signal. *This is addressed in Chapter 2*.

2. **What are the main variables influencing the characteristics of the leak noise?** Experimentally quantify the factors influencing a leak signal, including: leak flow rate, leak shape, leak area, system pressure, backfill type and pipe material. *The methodology section in Chapter 3 describes the experimental procedure and the results for this objective are addressed in Chapter 4*.

3. **What are the characteristics of the leak noise for more representative leak types/shapes found in real WDS?** Experimentally explore the influence of leaking electrofusion joints on the leak signal, which are more representative of 'real' leaks on plastic WDS. *The methodology section in Chapter 3 describes the experimental procedure and the results for this objective are addressed in Chapter 4*.

4. **What is the current state of research knowledge that will enable the development of leak flow rate, shape and area classification model?** Evaluate and explore different methods of leak signal classification or prediction, in order to classify/predict leak flow rate, allowing water companies to prioritise leak repair based on leak flow rate. *This is addressed in Chapter 5*.

5. **Is it possible to predict leak flow rate, shape and area from leak VAE signals?** Design a classification/predictive tool for leak area and leak shape in order to enable water company practitioners to rapidly prioritise leak repair based on leak area and leak shape. *This is addressed in Chapter 6*.

Chapter 2

Literature Review

2.1 Leakage in water distribution systems

2.1.1 Introduction

Water Distribution Systems (WDS) play an important role in safe guarding human health. A safe and reliable method of transporting drinking water is crucial to maintaining this. Typically, 20-30% of water pumped into the pipe network is lost through leakage, but this can be as high as 50% in developing countries and in locations where there are older distribution networks (AWWA, 1987; Almeida et al., 2014a). This represents a substantial loss of revenue and energy, as pumping and treating water has been reported to use between 2-3% of the worlds energy consumption (Almeida et al., 2015b). In the UK, leakage alone has been estimated to cost the government £7bn annually in street works costs (Almeida, 2013), as well as further social and damage costs (McMahon et al., 2005) and negative environmental and economic effects (Colombo and Karney, 2002). Leakage has become particularly important due to the loss of revenue, rising customer water bills, population growth and environmental concerns. Leaking pipes also pose a risks to public health due to risks to water quality (Fox et al., 2014), through the potential introduction of contaminants external to the pipe (e.g. fertilisers, pesticides, solvents etc.) (Lechevallier et al., 2003) through contaminant ingress.

The definition of water loss is complex, but is generally referred to as water that cannot be accounted for, i.e. Non-Revenue Water (NRW). NRW consists of real losses which are the real, physical water losses from the supply network, and apparent losses, which are illegal or unauthorised losses in the network (such as water theft) and metering inaccuracy. A traditional water balance does not determine where the losses in the network occur; moreover it only suggests that water loss in the network is occurring within a particular District Metered Area (DMA). Some of these losses are difficult for a water company to control. Water loss due to leaking pipes (real losses) are a waste and there is strong regulatory pressures from Ofwat to drive down leakage rates. Therefore, water companies aim to reduce the volume of water lost. In addition to regulatory pressures, reducing leakage provides major benefits to the water company, in terms of economic savings and in terms of safeguarding water quality (if there are no leaks in the water supply system the potential for contamination is reduced).

Usually in the UK, WDS are divided into transmission pipelines consisting of pipe >300 mm, and distribution mains which contain water distributed to industrial or residential areas (Pal, 2008). The extent of each pipe material varies depending on location, with most of the UK WDS made up of Cast Iron (CI), Ductile Iron (DI), Asbestos cement (AC), Copper (C), Polyethylene (Low Density Polyethylene (LDPE)/Medium Density Polyethylene (MDPE)/High Density Polyethylene (HDPE), Glass reinforced plastic, concrete and some Lead pipes still exist. Within the UK WDS, CI pipe represents the majority of the material at 40-60% (Pal et al., 2010). However, many of these have high failure rates due to corrosion and low strength, especially in older distribution systems. Most new pipes are laid with plastic pipe material as they are comparatively stronger and have a greater degree of flexibility, and therefore have greater resistance to the causes of bursts and leaks (such as hydraulic transients and ground movement). Plastic pipes are also less likely to suffer from corrosion. It has been assumed that plastic pipe are leak free (Fox, 2016), hence in part this is the reason for their popularity. However, plastic pipes are far from leak free, as plastic pipes are particularly prone to leaking at joints (Tayefi, 2014). The shear variety of pipe materials and other physical variables that come in to contact with a pipe means that leaks in the WDS take many forms. Many are formed through structural failure, giving rise to leaks of different shapes and sizes. The failure mechanism is specific to the manufacturing process, method of installation, loading and environmental conditions and pipe material (Fox, 2016).

Leaks in the water supply network can be identified using a number of techniques. However, the efficacy of many methods is questionable, with some techniques particularly limited by pipe material. Hence water companies, often working in collaboration with industrial research and development and academia, are continuously exploring new and innovative methods of detecting leaks in WDS. Leak detection in pipelines is a well-reviewed subject and the available literature covers a variety of techniques ranging from managing the whole WDS, to methods which asses leakage on individual sections of pipe (reviewed by Cist and Schutz (2001)), i.e. at the asset scale. Puust et al. (2010) provides a succinct review of leak detection techniques in combination with leakage management strategies.

Largely, at the DMA scale, leaks are managed through careful pressure management. However, a statutory minimum required pressure of 10 m at the customer tap through Ofwat's Guaranteed Standards of Service scheme (ThamesWater, 2018), enforces that pressure cannot be continuously reduced in order to fully eliminate leakage. At asset level scales, some methods to detect leaks include using pressure transients (Shucksmith et al., 2012; Ghazali et al., 2010; Ghazali, 2012), underground imagery (Ground Penetrating Radar (GPR) and thermography) (Demirci et al., 2012; Ayala-Cabrera et al., 2013), chemical and gas tracer methods (Suez, 2015), Vibro-Acoustic Emission (VAE) methods (Gao et al., 2006; Almeida et al., 2014a; Pal, 2008) and online monitoring methods (Sadeghioon et al., 2014). More recently, there have been advances in the use of satellite imagery (Severn Trent, 2017) and drones to detect leaks (Water, 2017), which may be termed as 'disruptive innovation' due to their potential in locating a large number of leaks in a short space of time. The efficacy of all of these methods can differ depending on the characteristics of the leak (e.g. leak shape, size etc.), pipe material and backfill, amongst other factors (Pal, 2008). Despite a wide variety of tools and techniques available, detecting leaks via VAE methods whereby a 'leak noise' is used to find the location of the leak is still the most common practice to detect leaks in the UK. Therefore, the research presented herein concentrates solely on VAE methods. Other techniques are not reviewed here and for a full and succinct review of many of these alternate techniques, see Puust et al. (2010).

2.1.2 Sustainable Economic Level of Leakage

Water companies are required to meet leakage targets set by Ofwat and meet their Sustainable Economic Level of Leakage (SELL), whereby sometimes it is cheaper to leave a leak unrepaired rather than undertake costly leak repairs. The SELL mechanism represents a trade-off between the economic cost of fixing the leak and the cost of water lost, and considers further options such as alternative methods of supply. Ofwat considers that the SELL mechanism offers customers the "best value for money" (UtilityWeek, 2018), and has published guidance on how social and environmental issues can be incorporated into the SELL assessment (RPS, 2008). However, the use of SELL does not consider the potential of bacteriological contamination due to negative pressure in the network, and therefore by leaving pipes to leak the risk to water quality is increased through contaminant ingress (Fox et al., 2014). There is also little known about how the shape of the leak or the changes to the leak area with time (i.e. the leak area may grow), and therefore there may be a point when a decision made to leave a leak under the SELL assessment no longer represents the best economic, social and environmental option. Moreover, leaks of higher leak flow rates represent a greater loss of water from the distribution network and therefore prioritisation of repair to higher flow rates can reduce SELL by making it more economic to fix less leaks. However, in catchment areas such as London, which is highly urban, repairing every leak can cause major disruption and cost, the SELL concept is therefore useful.

2.2 The leak noise

As water discharges through a leak, it generates a noise/vibration which travels down the pipe wall and in the fluid. This leak noise can be used in order to detect the presence and location of a leak. The aim of this section is to briefly review some aspects of the leak noise, including how the noise is generated, the detection of leaks using the leak noise and the factors that affect the leak noise which can also influence the efficacy of leak detection using VAE techniques.

2.2.1 Generation of the leak noise

The fundamental form in which the leak noise is generated has received minimal investigation in the research community, whereby the existing small body of literature assumes that the generation of the leak noise is due to either cavitation or turbulence. Cavitation is the formation and collapse of small bubbles due to changes in pressure, whilst turbulence occurs as the water discharges through the leak hole/crack. A study by Ruchonnet et al. (2012) suggested cavitation was very active in the generation of noise in nuclear pipes. Qing et al. (2006) suggested that the cause of vibration in pipes was due to cavitation from an orifice. In an idealised cavitation bubble, a bubble will develop, contract, detract and contract again. However, bubbles in real systems tend to break up into microbubbles when they collapse due to the shock wave generated during collapse. Therefore, any microbubbles observed due to leaks may be following bubble collapse. Studies in other disciplines suggest that, should cavitation occur, sound can be generated at a wide range of frequencies (Ross, 1976) which would likely be detected by downstream accelerometers and/or

hydrophones. It is therefore highly possible that cavitation plays a role in the generation of the leak noise.

Turbulence is a classical problem in fluid dynamics (Reethof, 1978) and is known to produce noise within in-pipe flows. Turbulent flows through a leak opening are also potential causes of the leak noise, due to friction between the discharging water and the pipe material as the leak is forced from an area of high pressure to low pressure. This will depend on the roughness of the leak at the solid-liquid interface, as well as the shape and size of the leak and the system pressure. This turbulence can cause vibrational and acoustic energy which transmits along the pipe wall and fluid, and therefore can be measured using sensors to find the leak. Thompson et al. (2001) states that the lower frequency bands found within a leaks acoustic emission signal are due to turbulence and the higher frequencies are due to cavitation. However, Thompson et al. (2001) does not include any actual data or figures within the study or any insight in to the work conducted. A study by Papastefanou (2011) provides the only known peer-reviewed published study, and also suggests similar conclusions by monitoring bubble formation in clear plastic pipe. They concluded that cavitation was not responsible for the leak noise as bubble formation was not observed, suggesting that if microbubbles are present that cannot be seen, they are related to higher frequencies. Assuming that smaller, non-visible microbubbles are responsible for high frequency sound, it is unlikely that cavitation would be the cause of the leak noise considering the leak noise is usually located in lower frequency bands in plastic pipe (Muggleton and Yan, 2013). Moreover, due to the rapid attenuation of high frequency signals in plastic pipe (Almeida et al., 2014a), even if cavitation occurs it may not be recorded by the accelerometers.

Whilst the existing body of literature appears to suggest that cavitation or turbulence is responsible for the generation of the leak noise, there appears to be no consensus in this understanding. This highlights a major research gap in the understanding of the fundamental mechanisms generating the leak noise, as whilst it is likely due to cavitation or turbulence, it may also be combination of both. A better understanding of the generation of the leak noise may help to develop understanding of leak noise transmission on plastic pipelines and ultimately how effective these leaks can be detected using VAE based techniques.

2.2.2 Leak detection using the leak noise

Leak detection via VAE based methods is common due to its relatively high level of accuracy and ease of use. Leak detection via VAE methods takes many forms. Traditionally, a DMA is assessed by an experienced leak detection operative (or 'Leakage Technician') using a listening stick (Pal, 2008), where a metallic rod with diaphragm is placed in contact with the pipe at convenient access points (e.g. valves and fittings). The Leakage Technician then listens to the VAE signal transmitting through the listening stick towards the Technicians ear (Pal, 2008). However, whilst this may detect the presence of a leak somewhere along the pipe line, it does not provide any estimate of the leak location and therefore other tools are required in order to fully localise the position of the leak in the pipeline.

A more advanced method of using leak VAE signals is that of leak detection and localisation via leak noise correlation, which has been a common method for over 30 years (Fuchs and Riehle, 1991). As water is discharged through a hole/crack in a pipe, an acoustic sound is created and

propagates in the fluid and a vibration propagates along the pipe wall, which can be recorded by accelerometers or hydrophones placed on or in access points (such as valves or fire hydrants). An example of a typical set-up using accelerometers or hydrophones for leak noise correlation is shown in Figure 2.1. Here, sensors are used to record the leak's VAE to determine the its location. Leak noise correlation is then performed by cross correlating the leak noise created as water discharges from the leak. Traditionally, cross correlation can be applied to applications where the data set is shifted in time or space, in order to find the time lag between two signals.

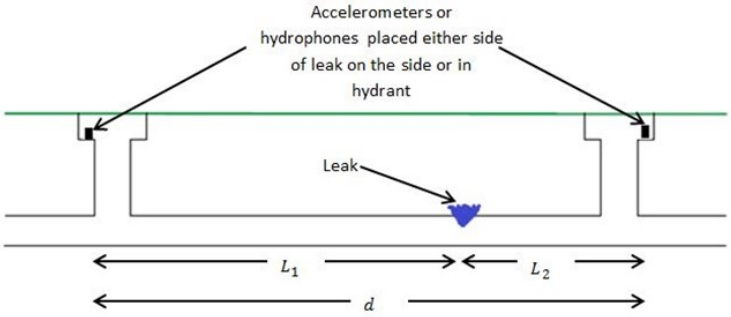


Figure 2.1: *Schematic representing the deployment of accelerometers or hydrophones to record the leak noise and localise the position of the leak.*

When sensors are placed either side of the leak and the leak noise is cross correlated, the leak's location is given by:

$$L_2 = \frac{d - c\tau_{delay}}{2}, \tag{2.1}$$

where c is the wavespeed of the leak noise on the pipe wall and d is the total distance between the two accelerometers or hydrophones. τ_{delay} is the arrival time of the leak noise and is estimated by cross correlating leak signals recorded by sensors deployed either side of the leak. Both the leak signal (S) and the background noise $(n1(t)$ and $(n2(t))$ can be modelled for signals recorded by accelerometer 1 (x_1) and accelerometer 2 (x_2) as follows:

$$\begin{bmatrix} x_1(t) = S(t - \tau_1) + n_1(t) \\ x_2(t) = S(t - \tau_2) + n_2(t) \end{bmatrix} \tag{2.2}$$

where τ_1 and τ_2 represent a time difference equivalent to the time difference of arrival between the received signal. These leak signals travel up and down the pipe in both directions, and become rapidly attenuated when moving further from the leak. Assuming that the background noise is uncorrelated with each other and uncorrelated with the leak signal, the Basic Cross Correlation (BCC) of the signals is described by:

$$R_{x_1x_2} = E[x_1(t)x_2(t + \tau_{delay})]. \tag{2.3}$$

where $E[]$ is the expectation operator. The BCC can take place in both the time and frequency domains, with the latter known for being more computationally efficient. It is common to express the cross correlation in its normalised form (-1 to +1), providing a cross correlation coefficient given by Almeida et al. (2014a) as:

$$\rho(\tau) = \frac{R_{x_1x_2}(\tau)}{\sqrt{R_{x_1x_1}(0)R_{x_2x_2}(0)}}, \tag{2.4}$$

where $R_{x_1x_1}$ and $R_{x_2x_2}$ are autocorrelation functions at sensor position 1 and 2 respectively when $\tau = 0$.

Ideally, the cross correlation of a leak signal received at the two sensors ($x_1(t)$) and ($x_2(t)$) produces a distinct peak to provide an accurate time delay estimate, when two signals are delayed from each other in time. An example of a good correlation between a leak signal recorded by 2 accelerometers placed either side of the leak is given in Figure 2.2. Here, a distinct, clear peak is demonstrated within the cross correlation. However, a number of factors can influence the leak signal which can create additional peaks in the cross correlation (such as excessive background noise), affecting the accuracy of leak location estimates. The accuracy of the time delay estimate also depends on a good signal to noise ratio (SNR). The addition of multiple broad peaks in the cross correlation also result in inaccurate time delays, which can lead to dry holes.

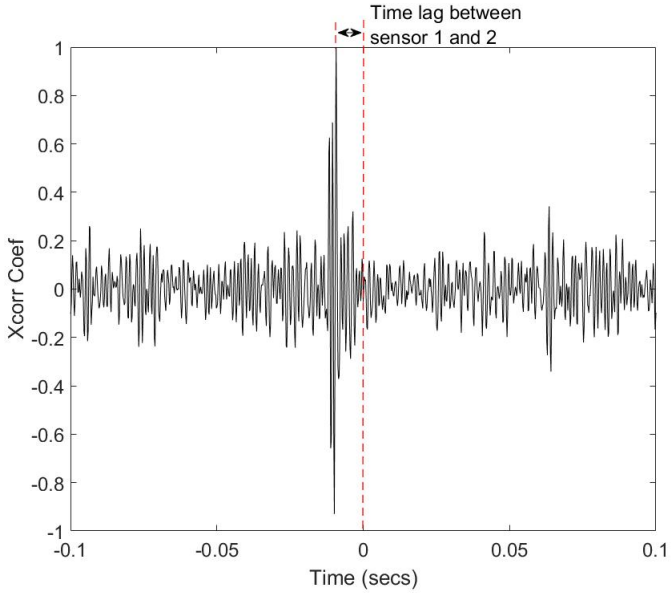


Figure 2.2: *An example of a good quality correlation between two leak signals.*

Key to the leak location estimate (Eq. 2.1) is the estimated wavespeed (c) of the leak signal, which is known to vary due to pipe material, diameter, wall thickness and the presence of discontinuities within the pipeline (Pal, 2008). As the wavespeed is required to be known *a priori*, standard equations are normally used (Almeida et al., 2015a) by Leakage Technicians using look up tables (Pal, 2008). The wavespeed is given by Almeida et al. (2014a) as:

$$c = \frac{c_f}{(1 + \frac{2Ba}{Eh}^{0.5})} \tag{2.5}$$

where c is the wavespeed, c_f is the free-fluid wavespeed, B is the fluid Bulk Density, E is the Young's modulus and a and h are the pipe radius and the thickness of the pipe wall respectively. Commonly, Leakage Technicians select the wavespeed manually from theoretical equations relating to the pipe material and diameter, but in actual measurements of wavespeed this can vary from case to case. For example, Muggleton et al. (2006) found wavespeeds of 160 m/s in MDPE pipe, whilst Muggleton and Brennan (2008) found wavespeeds of 365 m/s in the same pipe material. Whilst reports of wavespeed differ for plastic pipe, it appears that they are commonly reported as

propagating at a much slower speed than metallic pipes (Brennan et al., 2006a), and are normally estimated to be between 150 m/s and 400 m/s. The varying reports of the wavespeed for plastic pipes suggests that the use of theoretical equations may lead to error in the leak location estimate. The calculation of an inaccurate wavespeed can result in inaccurate leak locations, leading to a 'dry hole' (Almeida et al., 2015a). Due to the uncertainty surrounding wavespeed estimates, a variety of other methods have been proposed to estimate the wavespeed, including from the phase-frequency relationship (Gao et al., 2002), pipe excitation (Almeida et al., 2015a) and correlation method (Pal, 2008). However, these are less commonly used by Leakage Technicians who usually favour using the theoretical equations (Almeida et al., 2014a).

There are a number of wave modes within fluid-filled pipes, but the wave associated with the leak noise is carried within the fluid (Brennan et al., 2006b). Moreover, there is a degree of uncertainty with regards to the wavespeed in plastic pipe due to the wide-ranging variables that can influence propagation speed. Brennan et al. (2006b) carried out physical investigation in the factors influence the wavespeed, finding that in addition to the pipe thickness, pipe diameter, material and temperature all influence propagation speed. However, the backfill had little influence on the wavespeed up to approximately 500 Hz.

2.2.3 Variables effecting the characteristics of the leak noise

A number of variables have been reported in the literature which affect the characteristics of the leak noise. The variables can effect the leak signal at source, but some variables are more transmission dependent, influencing the characteristics of the leak signal as it propagates along the pipeline to the sensors. This section aims to review the factors influencing the leak noise.

2.2.3.1 Relationship between leak flow rate and leak vibro-acoustic emission

Increasing WDS pressure is well known to cause an increase in leak flow rate (Miller et al., 1999). Traditionally, leak flow rate (Q) has been shown to be sensitive to pressure through the orifice equation (Greyvenstein and Zyle, 2004):

$$Q = C_d A \sqrt{2g\nu} \tag{2.6}$$

where g is acceleration due to gravity, C_d is the discharge coefficient, A is hole area, pressure ν and Q is the flow rate through the leak. Therefore, the above equation is often used in pressure management in order to calculate the level of leakage reduction alongside a reduction in pressure. Numerous studies have investigated the applicability of the orifice equation to leaks in WDS, resulting in the development of the power equation which is the preferred method to model this relationship (Cassa and van Zyl, 2011):

$$Q = \varpi^{\alpha} \tag{2.7}$$

where α is the leakage exponent and ϖ is a leakage coefficient. α can vary due to leak hydraulics, pipe material, surrounding media and pressure (Greyvenstein and Zyle, 2004).

The increase in leak flow rate has in turn shown to increase the amplitude of the leak's VAE

signal in numerous studies (Hunaidi and Chu, 1999; Pal, 2008). This agrees with theory that for fixed sized leaks, higher pressure results in a higher leak signal amplitude due to increased leak flow rate (Ver, Istvan, Beranek, 2005). Similarly, Papastefanou (2011) and Pal (2008) demonstrated increasing signal amplitude with increasing pressure due to the strong influence of leak flow rate. Pal (2008) also found leak flow rate increased leak VAE frequency. Papastefanou (2011) established an empirical relationship between leak size, amplitude and leak flow rate and continued to suggest that it is easier to detect leaks of a higher flow rate compared to those at lower flow rates. It was also noted by Humphrey et al. (2012) that the increased amplitude of leak VAE signals with increased leak flow rate may make it easier to detect a leak through a more pronounced and defined peak in the cross correlation. The study investigated the influence of leak flow rate on correlation performance, finding that leaks with flow rates of 0.5 m^3/hr at a distance of 186 m from the leak had a low success rate in detection, whereas leaks at higher flow rates of 1 m^3/hr at the same distance were detected more successfully. However, increasing the leak flow rate to 1.5 m^3/hr and increasing the measurement distance to 316 m did not produce any successful correlations.

Evidently, the leak flow rate has a strong effect on the leak signal. Moreover, the leak flow rate is linked to the SELL assessment, in that it may be more cost-effective not to repair leaks with small leak flow rates as the cost of repair is greater than the cost of water lost. Knowledge of leak flow rate could allow water companies to prioritise leak repair activities and save more water by fixing less leaks. An attempt was made in order to use leak VAE to estimated the flow rate of the leak, known as the 'Leak Sizer'. The Leak Sizer categorizes leaks within a range depending on their flow rate (Primayer, 2016). It was claimed by the manufacturers that 83% of water can be saved by repairing only 40% of leaks. However, the Leak Sizer was not actively used by water companies. There are also no studies (to the knowledge of this author) investigating the accuracy of this technique and due to the commercial value of the Leak Sizer, and little information exists as to the signal processing methods used. Other academic led studies have attempted to quantify leak flow rate in gas pipes (see for example Chen et al. (2007), Kaewwaewnoi et al. (2010), Kaewwaewnoi et al. (2007), Meland et al. (2011b)). Despite some study in academia and the availability of the Leak Sizer, there exists no other method to estimate the flow of the leak from VAE signals in WDS that is known to the author of this research. However, there is a growing industry need to find ways to further drive down leakage levels.

2.2.3.2 Relationship between backfill and vibro-acoustic emission

The process of leak repair under usual circumstances involves excavating the ground surrounding a suspected leak location, repairing a leak and then refilling the excavated hole. The pipe itself then becomes supported by the surrounding backfill, and the type of backfill can vary depending on geographic location. The influence of the type of backfill has been investigated by only a few authors, but is often cited as a major factor influencing a leak signal (Muggleton and Brennan, 2004; Pal, 2008). The leak signal is likely to be strongly affected by the backfill as the leak discharge pattern is influenced by and interacts with the surrounding media, where some of the leak energy can be absorbed by surrounding media particles (Pal, 2008). Van Zyl et al. (2013) reported that fluidisation of the surrounding media can occur in backfill surrounding a leaky pipe, which can result in the mobilization of backfill particles when agitated by the leak. This in turn can create a sound as media particles mobilise by hitting the pipe wall, which is likely to interact with the leak signal. The surrounding media has also been reported to influence the attenua-

tion of signals (Brennan et al., 2006a), and can be due to the levels of compaction, soil type and degree of saturation and changes to the soils structure such as eroding the ground (Reed et al., 1989).

The influence of soil type and temperature on leak acoustics was investigated by Hunaidi and Chu (1999). The investigation took place on a buried test facility in Ottawa Canada, where frost can penetrate to depths of 1.5 m. They found that at frequencies <10 Hz, signals were similar in summer and winter, although there appeared to be a slightly higher attenuation rate of the lower frequency bands during the winter. The main frequency peak was found to move from 65 Hz to 55 Hz in the summer. The reason for this shift in peak is unknown, however this may be due to a variation in the coupling between the soil and the hydrant (where the signals where measured) due to freezing (Hunaidi and Chu, 1999). Brennan et al. (2006a) also analysed the influence of the surrounding media by burying a pipe in soil and air, reporting that there was no significant effect of the surrounding media on frequencies up to 500 Hz. However, they reported decreasing amplitudes away from the leak with increasing frequency on the buried pipe compared to a pipe in air, likely due to the transmission of energy into the ground. The main losses in this study appeared to be due to material damping in the pipe wall at frequencies up to 100 Hz. Brennan et al. (2006a) went on to bury the pipe in water, and found that there was a slight decrease in wavespeed and a small increase in attenuation. Moreover, Muggleton et al. (2002) and Fuller and Fahy (1982) demonstrated that the surrounding media had little influence on the axisymmetric wavenumber.

A study by Muggleton and Brennan (2004) detailed the importance of understanding surrounding media and made comparisons between submerged pipes and pipes buried in soil, finding that submerged pipes attenuated less compared to that of pipes buried in soil. Although this study provided interesting results, it was limited by the fact that comparisons were made between two separate pipe rigs, and therefore the conditions could not be compared as the variables and conditions are likely different (including the leak geometry, leak flow rate etc.) when switching between the experimental sites. Pal (2008) also investigated leaks under sandy soil compared with leaks discharging to gravel, finding that there was no significant effect of backfill on the leak frequency response. However, in this study, there was no attempt to control other variables such as leak flow rate, which is likely to change depending on the backfill type and degree of compaction. Therefore, there appears to be a significant research gap; all previous studies that have attempted to experimentally quantify the effect of backfill on the leak signal have made no attempts to isolate or control the other variables (such as leak flow rate) which can contribute to a change in leak signal. These studies are therefore limited by this fact meaning that the conclusions cannot be verified. The influence of backfill therefore does not appear to have been experimentally determined. The isolation of other variables which have been shown to influence the leak signal is the first step in quantifying the effect of backfill.

2.2.3.3 Relationship between leak area, leak shape and vibro-acoustic emission

A variety of leak shapes and sizes can occur in WDS, including pin holes, joint leaks, circumferential cracks and longitudinal cracks (UKWIR, 2008), amongst others. It is known that some leak shapes are more likely to increase leak area compared to others (Fox, 2016). Experimental studies investigating the hydraulic nature of differing leak shapes in plastic pipe have shown that longitudinal slits have a time-pressure dependent leak area, with a growth in leak area with increased pressure (Ferrante, 2012) and a hysteric nature. However, round holes have been reported as less likely to

grow with pressure (Fox, 2016). Leaks commonly studied on plastic pipe in terms of VAE signal experiments are normally one of three shapes, including round holes (Pal, 2008; Wassaf et al., 1985), artificial leaks from fire hydrants (Gao et al., 2005; Almeida et al., 2014a) and slits (Pal, 2008; Wassaf et al., 1985). Despite the variety of leak shapes existing in pipelines, there has been limited study investigating the effect of leak shape on the VAE signal. Brunner and Barbezat (2007) used screws of different hole diameters to simulate round hole leaks of varying diameters, finding differences between the power spectra of the different diameter leaks. Wassaf et al. (1985) compared the VAE response of leak shapes of circular holes and rectangular slits, showing that both shapes had a different leak frequency spectra. Pal (2008) attempted to compare the VAE signal of different leak shapes, assessing artificial leaks created by fire hydrants, split leaks and joint leaks made by loosening the nuts on a flange plate. Pal (2008) found that the frequency spectrum of a leak was strongly governed by its leak shape and size. However, although these studies provided an interesting perspective, the leak flow rates were not controlled when making comparisons between any of the leak shapes investigated in any of these studies. As leak flow rate has been shown to have such a strong influence on the leak signal (Pal, 2008), any assessment of leak shape requires a good experimental methodology which controls leak flow rate between shapes. Therefore leak shape is isolated and it's effect on the leak signal can be understood. This is not the case in the previous studies.

Although the aforementioned studies provide a useful insight in to the effect of leak shape on the VAE signal, the leak shapes normally studied are uncommon in plastic pipes on real WDS (Water Services Association of Australia, 2012) and therefore have little representation of real leaks that would be found in a real WDS. The majority of leaks in plastic pipes actually occur due to leaky joints (Tayefi, 2014; Water Services Association of Australia, 2012), typically due to contamination of the joint when the pipe is being installed or repairs made. This is especially true in the case of electrofusion and buttfusion weld joints (Tayefi, 2014). However, there have been no studies investigating the VAE signal produced by leaks in leaky electrofusion or buttfusion weld joints in a controlled environment. Therefore, an assessment of leak shape with the inclusion of leak shapes that are representative of real leaks represents a major research gap as 'real leaks' are seldom investigated.

The influence of leak area on leak signals has undergone some attention in the literature, especially in terms of leak hydraulics. Leak area is a key parameter in the pressure-flow relationship (Equation 2.6), and is therefore key in defining the volume of leakage reduction using pressure management. A study by Fox et al. (2014) also showed that there was potential for contaminant ingress through leaks during periods of negative pressure. Therefore, a larger opening due to increased leak area size increases the risk of contaminant ingress. Many studies claim to have assessed the influence of leak area on leak VAE signals (e.g. Sun et al. (2016) and Meland et al. (2011a)). However, in these studies system pressure was kept consistent for a change in leak area, and therefore the changes in leak signal are more likely related to changes in leak flow rate [due to increased leak area], where often authors mistakenly suggest the change in signal is due to the variation in leak area. The only way to understand the effect of leak area is by comparing leaks of varying area and controlling the leak flow rate such that the effect of leak area is isolated from all other variables, especially leak flow rate.

An interesting study by Sun et al. (2016) showed that the area of round hole leaks in gas pipes could be classified using VAE in combination with Support Vector Machines. The ability to estimate

leak area or leak shapes non-invasively has never been undertaken in the context of leaks in the water industry, although it could be of great benefit. Any system that allowed for the classification of leak area and leak shape may allow water company practitioners to prioritise leakage repair strategies by repairing leaks of larger area and therefore of greater risk to contaminant ingress and shapes more likely to grow. Despite the promise shown in gas pipes by Sun et al. (2016) a method to classify leak area and leak shapes from VAE in WDS does not exist.

2.2.3.4 Transmission effects on the leak noise

Often, WDS are divided into transmission pipelines consisting of pipelines >300 mm in diameter, and larger diameter distribution mains which contain water feeding DMAs. The extent of each pipe material varies depending on location, with the majority of the UK WDS made up of Cast Iron (CI), Ductile Iron, Asbestos cement (AC), Copper, Polyethylene (PE), Glass reinforced plastic, concrete and some lead mains still exist. Within the UK WDS, CI pipe is the most common, consisting of approximately at 40-60% of the pipe material (Pal, 2008). Many of these materials have high failure rates due to corrosion and low strength, especially in older distribution systems. Nowadays, most new mains are laid with plastic pipe as they are comparatively stronger and have a greater degree of flexibility. Therefore, they are better at resisting the causes of bursts (such as hydraulic transients and ground movement) and are also less likely to suffer from corrosion.

Pipe material has been shown to influence the hydraulic behaviour of leaks. It was found by Greyvenstein and Zyle (2004) that the pipe material has an influence on the leakage exponent. Moreover, Ferrante (2012) found different head discharge relationships for different materials. Pipe material has also been shown to be a important factor in influencing the characteristics of the leak noise. Leak noise correlation is known to be relatively successful on metallic pipes, but there is often difficulty when detecting leaks in plastic pipes (Almeida et al., 2014a), with many studies reporting varying efficacy (Pal, 2008; Hunaidi et al., 2004; Gao et al., 2006; Humphrey et al., 2012). Leak detection using noise correlators is often ineffective on plastic pipe due to the rapid attenuation of the leak signal (Almeida et al., 2014a) (Pal, 2008). This suggests that there is a strong influence of pipe material on the performance of leak noise correlators. The reduced efficacy of leak noise correlators on plastic pipe has also resulted in reduced confidence in using leak noise correlators by Leakage Technicians (Gao et al., 2002).

Due to the attenuative properties of plastic pipe, they are often referred to as low pass filters, in that they filter out some of the high frequency content of leak signals (Almeida et al., 2014b). Numerous authors have declared that leak signals in plastic pipe are low frequency, well below the pipe ring frequency (Muggleton and Brennan, 2004; Almeida et al., 2015b; Pal, 2008) and have a high degree of attenuation (Hunaidi et al., 2004). Some pipe materials, such as CI and AC are often discussed as having better transmissive properties compared to plastic pipe, however there are limited direct comparisons between pipe materials in the context of how they influence leak VAE signals. A study by Hunaidi et al. (2004) on a buried pipe rig in Canada suggested that leak signals travel further in metallic pipes. Mostafapour and Davoudi (2013) studied leakage of pressurised air from steel pipes, finding high bandwidth leak signals within a frequency range of 0 kHz and 300 kHz. Shehadeh et al. (2006) also studied steel pipes, measuring simulated leak noises of 120-180 kHz 5 m from the leak on short pipe sections. However, signal sources were simulated using an acoustic emission sensor and therefore do not represent a real leak signal. Pal (2008) investigated

the effects of different pipe materials on the leak signal and found a varying frequency range depending on pipe material, measuring frequencies between 10 to 750 Hz for AC, 10 to 800 Hz for Ductile Iron, 10 to 550 Hz MDPE without backfill and 10 to 250 Hz for MDPE with backfill. However, each leak had a different flow rate, and therefore it is not possible to state whether the frequency shift was due to a change in pipe material or leak flow rate.

Moreover, to improve the accuracy of leak noise correlation, leak noise correlators normally have set bandpass filters built in to them, and the set points depend on the pipe material. A report by Humphrey et al. (2012) suggests filters <100 Hz for plastic pipes and 1000-2000 Hz for CI, steel and AC pipes <150 m length. The influence of pipe material on leak VAE signals does seem to be recognised as Leakage Technicians commonly vary bandpass filter set points adapted to the different pipe materials. In real WDS, the propagation of acoustic/vibrational energy is dampened due to signal attenuation and radiation into the surrounding media. Moreover, the rate of signal attenuation is linked to the pipe material (Pal, 2008) with plastic pipe having much higher rates of attenuation than metallic pipes (Humphrey et al., 2012). Therefore, attenuation can have a significant effect on the performance of leak noise correlators. As sensors are moved further away from the leak, there will be a loss of leak energy making it harder to estimate the leak's location accurately. Transmission coefficients can be applied in order to demonstrate how the pipe will attenuate the leak signal and the attenuation of leak signals is given by Almeida et al. (2014a) as:

$$\delta = \frac{1}{c_f} \frac{n B a / E h}{[1 + (\frac{2 B a}{e h})]^{0.5}} \tag{2.8}$$

where δ is the attenuation factor and n is damping in the pipe wall. The actual attenuation (i.e. loss of sound per meter (dB/m)) is given by Gao et al. (2004) as:

$$Attenuation = \frac{20 Im\{k\}}{ln(10)} = -8.67 \delta \omega \tag{2.9}$$

This equation shows that higher attenuation occurs at higher frequencies, where a plastic pipe will effectively act as a low pass filter (Gao et al., 2004). Evidently, the pipe material appears to influence the transmission of the leak signal but this has not yet been experimentally quantified on real WDS, where the transmissive properties of different materials have been directly compared to one another.

Whilst pipe material appears to have a strong influence on the efficacy of leak noise correlators, some key limitations in the existing literature are apparent. The majority of leak detection research using leak noise correlators is conducted on pipe rigs or buried test facilities, which have little resemblance of the complexities of a real WDS. A real WDS is a complex system made up of a number of pipe materials with different diameters, valves, fittings, discontinuities and connections, which are likely to influence a leak signal along its transmission path. However, there are currently no studies comparing the transmission of the leak noise on plastic pipe in the context of other pipe materials with a good quality, well controlled experimental methodology. Moreover, some studies make comparisons within separate test rigs, whereby all the other parameters that can influence leak signals (including leak shape, flow rate etc.) are not controlled and therefore the effect of pipe material has not been isolated. Therefore, despite received wisdom, there is no experimental evidence investigating the influence of pipe materials on the leak signal with a good experimental

methodology whereby other factors that may influence a leak signal are standardised allowing for direct comparisons between pipe materials.

2.2.4 Enhancing leak detection via the leak noise through signal processing

In a real WDS, noise from traffic, valves, pipe elbows, and pipe resonance (Hunaidi and Chu, 1999) occurs to varying degrees. As mentioned, the basic cross correlation is commonly used in order to correlate the signals and provide the time delay (Equation 2.4). Ideally, the cross correlation produces a distinct peak when two signals are delayed from each other in time. However, a number of factors disturb the position of the peak, such as excessive background noise, pump noise etc. and the accuracy of the time delay estimate also depends on a good signal to noise ratio (SNR). However, in the presence of background noise the peak due to the leak is often suppressed. To improve leak location and thereby reduce 'dry holes' it is important these background noises are removed from the signal so the leak can be identified more clearly and accurately. Fortunately, cross correlation of leak signals has been shown by a number of authors to be improved through some additional signal processing techniques.

Typically, Leakage Technicians apply band pass filters to signals prior to performing cross correlation. This is often based on set filtering bands depending on the pipe material. However, the inherent problem with this is that in the field, Leakage Technicians do not know in advance the frequency of the leak and in some cases can actually filter out the leak signal (Hunaidi and Chu, 1999). However, it is possible to still apply suitable filters with prior knowledge of the pipe material: for example, plastic pipes tend to have leaks with lower frequencies (Almeida et al., 2014a), so a simple low pass filter can be used, but in the case of metallic pipes, the frequency of leaks ranges much further and therefore it is more difficult to select an appropriate filter. Moreover, the level of and power of background noise influences the SNR and thus strongly affects the accuracy of the time delay estimate (Ionel et al., 2010).

A number of filters are available but the most common method used in leak detection is that of bandpass filtering. Commonly, Butterworth filters are applied in order to remove background noise present in the signal (Singh and Singh, 2014), which has a flat frequency response in the passband (Taub and Schilling, 2002). This filter rolls towards zero and does not incorporate ripples in the passband. In terms of the transfer function ($H(\omega)$), the filters gain ($G(\omega)$) of the nth order low pass is:

$$G^2 = |H_{F(j\omega)}|^2 = \frac{G_0^2}{1 + (\frac{\omega}{\omega_c})^{2\vartheta}} \tag{2.10}$$

where G_0 is the DC gain and ω_c is the specified cut off frequency. Higher filter orders (ϑ) produce sharper cut off frequencies. Other filters, such as the Bessel filter provide a linear phase response and flat group delay, performing well at preserving the waveshape of signals in the passband. Due to the slow roll off of the Bessel filter, they normally require a higher order filter compared to the Butterworth filter.

It is well known that the presence of background noise can interfere with signals. This is especially important in the context of leak signals, whereby the repair involves major costs (due to streetworks, excavation costs etc.) and therefore the accurate localisation of leaks is of great

importance. A number of advances in the use of filtering have been shown to improve the location of leaks. Ionel et al. (2010) improved the quality of the cross correlation by filtering signals based on the signal bands with optimal coherence. Knapp and Carter (1976) introduced the method of Generalised Cross Correlation (GCC) which aimed to reduce these issues of background noise. The GCC introduces different weighting factors in the frequency domain, which provide a method of pre-filtering signals to enhance the SNR where the SNR is shown to be highest. Therefore frequencies outside these bands become suppressed (Gao et al., 2006). A variety of pre-whitening algorithms exist which can be applied in order to sharpen the peak in the cross correlation between leak signals (Gao et al., 2006). Gao et al. (2006) provided the most extensive comparison of these algorithms in the context of leak detection, whereby they compared the Maximum Likelihood (ML), Phase Transform (PHAT) and Smoothed Coherence Transform (SCOT) estimators to leak signals prior to cross correlation.

Initially, a weighting estimator (PHAT, SCOT, ML) are applied to the cross spectral density of the signal and the inverse Fourier transform is then taken. The SCOT estimator (Carter et al., 1973) weights the signal according to its SNR, attenuating the frequencies which have a low SNR and therefore those with a high SNR are accentuated. The SCOT transform was designed in order to identify the true time delay of signals with broadband noise at both sensors, and is given by:

$$\psi_{SCOT}(\omega) = \frac{1}{\sqrt{S_{x1x1}(\omega)(\omega)S_{x2x2}\omega)}} \tag{2.11}$$

The ML estimator (Knapp and Carter, 1976) was developed to determine the true time delay of two signals with uncorrelated noise. The ML estimator attenuates the noise power and accentuates the signal at the highest SNR, given by:

$$\psi_{ML}(\omega) = \frac{\gamma^2_{x1x2}(\omega)}{1-\gamma^2_{x1x2}(\omega)} \frac{1}{S_{x1x2}(\omega)} \tag{2.12}$$

The PHAT estimator solely computes the cross correlation based on signal phase. It has been noted that the PHAT performs well at high SNR, but deteriorates at lower SNR (Zhang et al., 2008):

$$\psi_{PHAT}(\omega) = \frac{1}{S_{x1x2}(\omega)} \tag{2.13}$$

$\psi_{PHAT}(\omega)$, $\psi_{SCOT}(\omega)$ and $\psi_{ML}(\omega)$ are then substituted into the GCC equation below in the frequency domain:

$$R^2_{x1x2}(\tau) = F^{-1}\{\psi(\omega)S_{x1x2}(\omega)\} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \psi_g(\omega)S_{x1x2}(\omega)e^{i\omega\tau}d\omega \tag{2.14}$$

where $\psi_g(\omega)$ represents the weighting functions of $\psi_{PHAT}(\omega)$, $\psi_{SCOT}(\omega)$ and $\psi_{ML}(\omega)$. F^{-1} is the Fourier transform and $S_{x1x2}(\omega)$ is the Cross Spectral Density, derived from:

$$S_{x1x2}(\omega) = \frac{1}{2\pi n_a T}\sum_{b=1}^{n_a} X^*_{1b}(\omega, T)X_{2b}(\omega, T) \tag{2.15}$$

where n_a represents averages of T.

The assessment by Gao et al. (2006) found that the application of the algorithms through the GCC process provided greater accuracy in the calculated time delay. In particular, the SCOT method was found to provide accurate time delays in comparison with the standard BCC estimator. However, Gao et al. (2006) analysed synthetic leak signal data along with a limited amount of experimental data. Therefore the ability of the estimators to perform when correlating leak signals affected by alternate variables (such as backfill, leak shape, pipe material etc.) has not yet been investigated. Moreover, there has been no assessment of how the estimators perform when correlating real leak signals from real complex WDS.

2.3 Models for leak flow rate, leak area and leak shape prediction

2.3.1 Introduction

The review of the literature so far has evaluated the influence of various variables on leak signals, identifying key research gaps in the current level of understanding. Of great interest is the influence of leak flow rate, area and shape. Whilst these parameters appear to have a significant influence on the leak signal, they are also parameters of high interest to water companies due to the interconnection between leak flow rate, area and shape with the the cost of the leaking non-revenue water. Knowledge of these parameters prior to excavating a leak could ultimately reduce SELL, by optimising the process by which leaks are repaired (i.e. by prioritising leak repair on the higher flow rate leaks and those which are more likely to grow). Although there has been little research on developing a method of predicting leak flow rate, leak area and leak shape in WDS, there has been some research focused on the estimations of leak flow rates through leaking gas pipe shut off valves and gas pipe leakage by Sun et al. (2016), Meland et al. (2011b) and Kaewwaewnoi et al. (2010). The methods to predict leak flow rate in these studies are usually based on one of two ways: (1) using single feature models focusing on data driven comparisons; and (2) multi-feature models, which are more in common with machine learning practices. The following section reviews a variety of modelling techniques that have been used in this context, or models which have the potential to be used in order to develop a leak flow rate, area and shape prediction tool.

2.3.2 Model selection

A number of model types can be used for the prediction of leak flow rate, leak area and leak shape, and these can be divided in to single and multiple feature models. A wide variety of single and multiple feature models exist, and they cannot all be reviewed or compared here. The models investigated below were chosen as they fit specific criteria (Meyers et al., 2017), which includes: (1) a model currently exists and therefore can be applied to the data collected; (2) proven to effectively evaluate non-linear and linear relationships with data sets with a potential degree of noise; and (3) they have direct applicability to the use of leak signals (e.g. many single feature models have already been proposed by Meland et al. (2011b) in a similar study with leaky gas pipes). As well as categorising the models in to single and multiple-feature models, predictive models can also be distinguished by model types, divided in to categorical (for classifying categories) or regression models. The estimation of leak flow rate is more suited to regression style models whilst leak shape is possible with classification based algorithms. Although there are a number of leak shapes in real WDS, they are best split in to broad categories. A report by the Water Service Association of

Austrialia (2003) documents the leak shapes likely occurring in water distribution pipes, which included corrosion holes, longitudinal slits and circumferential slits amongst others. Therefore, it is possible to divide leak shapes in to sets of categories, and therefore classification based tools are more useful in estimating the leak shape.

There are four key studies which relate to the research presented herein. This includes Sun et al. (2016) who used single feature models to predict the aperture of various diameter round holes in leaky gas pipes, Chen et al. (2007) who used single feature models to derive the leak flow rate through gas valves and Kaewwaewnoi et al. (2010) who estimated leak flow rate through leaky gas pipes also using single feature models. By the most extensive study is that of Meland et al. (2011b) who compares a variety of single feature and multiple feature models in order to estimate leak flow in gas pipes. Meland et al. (2011a) found the best performing model to be Kernel Partial Least Squares Regression (a non-linear alternative to Partial Least Squares Regression). However, Meland et al. (2011b) did not go on to compare their model to others which are more in common with machine learning practices, such as Least Squares Support Vector Machines which may have improved results further. Therefore, it is logical to begin with the single and multiple feature models utilised by Meland et al. (2011b), Sun et al. (2016) and Kaewwaewnoi et al. (2010) in order to identify whether or not leak flow, area and shape in WDS can be accurately estimated using these models or alternate models. The following section investigates the potential of single feature and multiple feature models for the prediction of leak flow rate, leak area and leak shape.

2.3.3 Single feature models

Single feature models are data driven, which involve the analysis of past leakage data to establish a model. Although this is not always the case, often a reference data set is created from historic data (Meland et al., 2011b).

2.3.3.1 Linear (and non-linear) regression

Regression methods model the relationship between two variables: (1) the inputs (regressors), x; and (2) the outputs (responses) y as described by Meland et al. (2011b). In the case of leak flow rate prediction, this would be x (signal value) and y (leak flow rate), creating a relationship in the form of:

$$y = xb + e \tag{2.16}$$

where the regression coefficient b reduces the residuals e. In the case of leak flow rate prediction, the value of x will be found and the regression coefficient b is found using a training data set (Meland et al., 2011b). The previously mentioned published studies from the gas industry (Kaewwaewnoi et al., 2010; Chen et al., 2007; Meland et al., 2011a; Sun et al., 2016) demonstrate a strong potential in using linear models, with good predictions of leak flow rate. The study by Chen et al. (2007) compared a number of signal parameters (or features) which were input in to a linear regression model, and they found that some features described leak flow rate effectively more than others. Kaewwaewnoi et al. (2010) also used linear regression methods, finding that this provided accurate results in determining leak flow rate. However, on the contrary the study by Meland et al. (2011b) found that the single feature models were poor predictors of leak flow rate

when compared to multiple feature models.

2.3.4 Multi-feature models

Single feature models tend to describe a single value within the leak signal and therefore can be vulnerable to changes, as if the single feature is influenced in some way then the whole model could be negatively impacted. A wide range of variables influence leak signals and therefore it is quite possible that a single feature alone cannot accurately predict leak flow rate, area and shape. Machine learning models can utilise multiple numbers of features and learn the mathematical model from the set of finite observations presented to it. The task of learning mathematical models for leak flow rate, leak area and leak shape prediction is a complex problem but could be achieved by a number of machine learning methods. However, due to the nature of this study, certain models are more suitable than others. For example, leak flow rate prediction in nature is qualitative and therefore predictive models are more suitable. Leak shape is given as a set of quantitative values (e.g. round holes, longitudinal slit etc.) whereby they are grouped in to a limited number of discrete categories, therefore classification models may be more appropriate. Thus, the approach of machine learning seems logical due to the complex nature of leak flow rate, area and shape prediction from VAE signals.

2.3.4.1 Multiple feature linear (and non-linear) regression

The regression model $y = xb + e$ can be adapted to include a number of features (two or more features), known as Multiple Linear Regression (MLR) and Multiple Non-Linear Regression (MN-LR), given by:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_{p-1} x_{i,p-1} + \epsilon_i. \tag{2.17}$$

whereby it is assumed that ϵ_i has a normal distribution with a mean of zero and the variance σ^2 is constant. β is the slope parameter, and $x_{i,1}, x_{i,2}....x_{p-1}$ represents the inclusion of multiple features in the model. Similar to single feature linear regression, this can easily adapted to include a non-linear kernel in order to fit non-linear data. Whilst there has been limited application of the MLR or MN-LR methods to leak detection or leak noise analysis, there have been a number of applications in other disciplines in the literature, see for example Lewis et al. (2015) and Mohamad et al. (2017).

2.3.4.2 Partial Least Squares Regression

Other multiple-feature models combined with Partial Least Squares (PLS) are discussed by Meland et al. (2011b), including Neural Network PLS, Spline PLS which are regression methods that use kernels to accept nonlinearities. The study by Meland et al. (2011b) found the best model for predicting leakage rates through valves was using a partial least squares regression (PLSR) with a non-linear kernel, thus termed kernel partial least squares regression (KPLSR). The objective of PLSR is to project predicted variables in to a new space and find the optimum hyperplane maximising the variance between the observed and predicted variables. Meland et al. (2011b)

suggests the Spline PLS and KPLSR are similar, and Neural network PLS is less robust and therefore were omitted from further analysis.

The basic model of PLSR is the formulation of an underlying linear model and is given as $y = xb + e$, where y is a $n * m$ matrix of responses, x is $n * p$ matrix $n * p$ of predictors, B is $p * m$ matrix of regression coefficients and E is an error (or noise) term (Statsoft, 2017). PLS produces factor scores, in that the predictor variables do not have any linear combinations (Statsoft, 2017).

PLS is known as an unrestrictive and flexible model, and is especially useful when the number of predictors outweighs the number of observations (Statsoft, 2017). PLS is conducted by extracting latent variables, which are not measured or observed but are an orthogonal set of factors (Khodayari-Rostamabad et al., 2009). Meland et al. (2011b) compared PLSR with the raw data, Principal Components Analysis (PCA) and Independent Components Analysis (ICA), whereby PCA or ICA are weighted sums of noise sources and therefore these are uncorrelated and statistically independent (Meland et al., 2011b). PLS with ICA mixing vectors was described by Meland et al. (2011b) as 'unstable' and they found the best results from using PLSR with non-linear kernels.

2.3.4.3 Logistic Regression

Logistic Regression (LR) is a popular method due to the simplicity of use and ease of interpretation. LR is a generalised linear model and is known for providing fast training times (Microsoft, 2017). LR determines the probability of occurrence through the fitting of log odds and explanatory variables to a linear model (Elio et al., 2017). In a situation when predicting the probability that a sample is class '1' rather than class '0', a learned function can be formed. The LR is model is given by Elio et al. (2017) as:

$$\log\left(\frac{P(Y=1|X)}{1-P(Y=1|X)}\right) = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n \tag{2.18}$$

where $X = (X_1 ... X_n)$ denote the explanatory variables, $Y = (0, 1)$ is the binary variable and $\beta = (\beta ..., \beta_n)$ denote the regression coefficients calculated from the data set. LR is based on the probability of occurrence, given by MedCalc (2017):

$$odds = \frac{ProbabilityOfCharacteristicsPresence}{ProbabilityOfCharacteristicsAbsence} \tag{2.19}$$

2.3.4.4 Decision trees

Decision trees are a method that is often used in machine learning and data mining, and can be utitlised for both categorical and continuous inputs. Based on a set of data inputs, the model develops a decision tree which is the learned function. They are often expressed by sorting the tree down from the root to a leaf node (Schmid, 1994). A tree represents a decision and is split depending on a differentiation of input variables. However, decision trees have been shown to be limited in real life conditions as it is difficult to collect a fully representative training set (Zorman et al., 1997), and a limited training data set can result in overfitting (Zhao, 2017).

Random Forests (RFs) are a type of ensemble learning that predicts from the ensemble to decision trees (Zhao, 2017). RFs have been shown to be an accurate and robust tool for many

machine learning tasks (Louppe, 2014). The RF algorithm divides the decision tree into numerous smaller independent classifiers and 1/3 of observations are left out of the fitting. The best split of data is then found by randomly selecting a subset of features and then looking for the best subset of those features. In the RF, a random set of attributes is selected for each node. Each node votes for a class and the resulting class is decided at the terminal tree which suggests the class depending on the class which received the maximum number of votes (Palczewska et al., 2013). The production of multiple decision trees from a random subspace improves generalisation accuracy (Hassan et al., 2015). Abdulsalam et al. (2007) state that RF model accuracy depends on a few factors, such as minimal correlation between the trees (i.e. minimal correlations result greater variance and therefore smaller error) and each tree strength (whereby stronger trees decrease error rates). So in essence, RF models are required to minimise the correlations but increase node strength (Breiman, 2001). As the RF averages several decision trees, the risk of overfitting is much lower and are usually favoured over decision trees due to the fact that they limit the amount of overfitting. Details of the process for fitting an RF model to data sets is given in Louppe (2014).

2.3.4.5 Least Squares Support Vector Machines

Support Vector Machines (SVM) are a method for solving classification problems with nonlinearity and pattern recognition (Shen et al., 2011). SVM has been used in leak detection in some rare cases, see for example Mashford et al. (2009). The objective of SVM is to derive the position of the optimum hyperplane which maximises the boundary between the training data. However, SVM models involve complex quadratic programming problems (Chelabi et al., 2016) and are known to take a long time for training (Shen et al., 2011). Least Squares SVM (LS-SVM) overcome the problems of SVM, whereby a number of linear equations are solved rather than quadratic equations (Suykens and Vandewalle, 1999; Shen et al., 2011). LS-SVM is therefore more optimal at solving non-linear systems, performing with higher accuracy (Si et al., 2016).

LS-SVM have been shown to be successful at predictive problems, such as identification of cracks in images (Gao and Yang, 2013), the characterisation of cracks in conductive materials (Chelabi et al., 2016) and faults in complex non-linear systems (Khawaja and Vachtsevanos, 2009). In the context of acoustic based measurements, Shen et al. (2011) used LS-SVM and wavelet packet decomposition of acoustic emission signals to classify the state of pressure vessels. There is no known use of LS-SVM in the context of leak detection known to the author.

The LS-SVM algorithm has been described by many authors (Chelabi et al., 2016; Si et al., 2016), following the theory by Suykens and Vandewalle (1999), and is described here for the completeness. $x_i \in R^n$ is data from a given sample $\{x_i, y_i\}_{i=1}^m$. Therefore, $y_i \in R^n$ becomes the output data. The classification is as follows:

$$\min_{w,b,k}^{J}(w,k) = \frac{1}{2}w^T w + \frac{1}{2}C\sum_{=1}^{m} k_i^2, \tag{2.20}$$

where J is the objective function, C is a regularization parameter (γ and σ), k is slack variable, b denotes bias and w is a weight vector. This equation (above) is subject to the equality constraint:

$$y_i = w^T\varphi(x_i) + b + k_i, i = 1, 2, ..., m, \tag{2.21}$$

where $\frac{1}{2}w^T w$ is a flatness measurement function, φ is a non-linear function mapping the input into a higher dimensional space which solves the regression problem (Si et al., 2016). A Lagrangian function can then be formulated:

$$L(w, b, k, a) = \frac{1}{2}w^T w + \frac{1}{2}C\sum_{i=1}^{m} k_i^2 - \sum_{i=1}^{m} \alpha_i[w^T\varphi(x_i) + b + k_i - y_i], \tag{2.22}$$

where α_i represents Lagrange multipliers. As the equality constraints follow Kuhn-Tucker conditions, the Lagrange multipliers can be positive or negative (Suykens and Vandewalle, 1999), given by:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^{m} \alpha_i \varphi_i(x_i) \\ \frac{\partial L}{\partial b}, = 0 \rightarrow w = \sum_{i=1} \alpha_i = 0 \\ \frac{\partial L}{\partial k_i} = 0 \rightarrow \alpha_i = Ck_i \\ \frac{\partial L}{\partial w} = 0 \rightarrow = w^T\varphi(x_i) + b + K_i - y_i = 0 \end{cases} \tag{2.23}$$

It is therefore possible to write a solution to a set of linear equations, $Ax = B$:

$$\begin{bmatrix} 0 & & I^T \\ 1 & K + C^{-1} & I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{2.24}$$

where $I = [1, 1, ..., 1]^T, A = [\alpha_1, \alpha_2, ..., \alpha_m]^T, Y = [y_1, y_2, ..., y_m]^T$. Mercer's condition is then applied to the matrix, observing the Kernel function as:

$$K(x_i, x_j) = \varphi(x_i)^T.\varphi(x_j) \tag{2.25}$$

It is now possible to derive the regression function for the LS-SVM algorithm:

$$f(x) = \sum_{i=1}^{m} \alpha_i K(x, x_i) + b \tag{2.26}$$

It is necessary to tune the hyper-parameters (γ and σ) in LS-SVM and this can be achieved using the leave one out method (Brabanter et al., 2011).

2.3.5 Model performance assessment

The model performance can be assessed through using the Root Mean Square Error (RMSE), which is a common parameter for assessing model performance (Gneiting, 2011). As the RMSE shows the residual error, it provides a good estimate of the difference between the LS-SVM model predicted values and the actual values (Zhan et al., 2015). RMSE is given by:

$$RMSE = \sum_{i=1}^{n} \frac{(x_i - \hat{y}_i)^2}{2} \tag{2.27}$$

where x_i is the real value of the test set, $\hat{y}_i$ is the LS-SVM output value of the ith sample and n is the number of samples.

2.3.6 Features to describe leak flow rate, leak area and leak shape

Of key importance of to the models described in the sections above is derivation of signal features. Normal features are just 'conventional variables' and regressors describing a specific part of the signal. Other regressors can be based on Independent Components Analysis (ICA) mixing vectors and frequency components. Meland et al. (2011b) compared all three as inputs in to the regression models PLS and KPLSR. When using ICA mixing vectors as regressors, Meland et al. (2011b) described the results as 'unstable', meaning that different RMSE values are obtained every time the model is trained. Therefore, only 'conventional' variables and frequency spectrum features are likely to be of use in the case of developing a leak flow rate, area and shape prediction model.

Conventional signal features can be be divided into categories, including analytical methods based on known relationships between parameters (such as the models described earlier) and data driven comparative methods such as pattern recognition (Meland et al., 2011a) and spectral comparison methods. As already stated, leak flow rate appears to increase signal amplitude (Pal, 2008). Similarly, an investigation into a leaking steam ball valve and water ball valve by Yan et al. (2015) demonstrated that signal amplitude is directly proportional to leakage rate. Khulief et al. (2012) found that it was easier to detect differences in sound power levels using signal Root Mean Square (RMS) when power levels are similar, when using acoustic emission of leaks in plastic pipes. Meland et al. (2011a) also found a good relationship between leak flow rate and RMS when investigating leaky shut down valves in the oil and gas industry. Moreover, Kaewwaewnoi et al. (2007), Kaewwaewnoi et al. (2010) and Chen et al. (2007) used VAE to investigate leak flow rate classification from gas leaks and hydraulic seals respectively, and both found good correlations between signal RMS and leak flow rate of a sample containing N samples, $x[0], x[1], ...[N-1]$:

$$RMS = \left(\frac{1}{2} \sum_{n=0}^{N-1} x[n]^2 \right)^{0.5} \tag{2.28}$$

The work by these authors shows that the VAE signal RMS can be related to the signal energy content (Chen et al., 2007). Chen et al. (2007) developed a leak flow rate prediction method using a simple least squares regression method based on taking the RMS of the recorded AE signal. Kaewwaewnoi et al. (2010) continued to develop an equation to relate VAE signal RMS from gas valve leakage to leak flow rate:

$$AE_{RMS} = C_1 \frac{1}{c^5 \rho^3 D^{14}} \left(\frac{Q}{C_V} \right)^8 \left(\frac{P_1 S}{ZP} \right)^4 \tag{2.29}$$

where AE_{RMS} is RMS value of the VAE signal. C_1 is a constant relating to valve material, signal attenuation and the effect of the sensor. ρ is the density of the fluid, D the size of the valve, Q volumetric flow rate, C_ν a valve flow coefficient, S is fluid specific gravity. C_1 is determined from experiments and ZP defines the pressure drop across the valve. However, this equation was evaluated by Meland et al. (2011b) and the model outputs were found to not be reproducible in their study and the leak flow rate predictions were high in error. Moreover, the model developed by Kaewwaewnoi et al. (2010) was for leaky gas pipes and therefore is likely to have little applicability to leaks in WDS.

Chen et al. (2007) compared signal RMS with several other several signal-energy related parameters to quantify gas leak rate, including VAE counts and the magnitude of the peak in the power spectral density (PSD). They determined the best feature for leak flow rate prediction from gas pipes by assessing the Pearson's Correlation Coefficient, which measures the linear association

between the feature and the leak flow rate (Hauke and Kossowski, 2011). The Pearson's Correlation Coefficient of variables x and y outputs a value of -1 to 1 depending on the degree and direction (positively or negatively correlated) of the correlation. The algorithm provides an indication as to the strength of the relationship between the variables (Zhou et al., 2016). The Pearson's Correlation Coefficient is given by Rodgers and Nicewander (1988) as:

$$r_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2}\sqrt{\Sigma(y_i - \bar{y})^2}} \tag{2.30}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i$ is the mean of x and $\bar{y} = \frac{1}{n}\sum_{i=1}^{N} y_i$ is the mean of y.

Chen et al. (2007) also compared RMS to VAE counts, whereby the number of times a sample exceeded a certain threshold is counted (for IEPE accelerometers this would usually be in volts). As leak flow rate is related to signal amplitude (Pal, 2008; Hunaidi and Chu, 1999; Papastefanou, 2011) it can be expected that higher leak flow rates result in a higher number of VAE counts. However, Chen et al. (2007) found difficulty using this feature as it's correlation with leak flow rate highly depended on the chosen threshold, and therefore it is difficult to make this effective for all leak flow rates when the physical conditions are likely to vary (which is the case for real WDS). Mba (2003) found good use of simple parameters such as VAE counts and RMS to detect defects in bearings within rotational machines. VAE counts are also a function of the sensor, damping characteristics of the material, signal amplitude and the chosen threshold level (Grosse and Ohtsud, 2008). As a result, VAE counts are less applicable to the wider water industry due to the variety of sensors used to record leak signals.

An alternate feature to those described above is to use frequency domain features. For example, the fundamental frequency has been shown by Prime and Shevitz (1996) to shift depending on whether or not a beam is cracked using vibration sensors. Other frequency domain methods such as the maximum and mean amplitude (dB) of a signals Power Spectral Density (PSD) was reported by Chen et al. (2007) as a good descriptor of pin hole leaks in gas pipes. The PSD is a common signal processing parameter that is used to compare all kinds of signals in the frequency domain. The leak signal PSD can be used to represent the power of signals over different frequencies. The PSD is defined by Marple (1987) as:

$$P[k] = \frac{T}{N}|X[k]^2| = \frac{T}{N}\left|\sum_{n=0}^{N-1} z[n]exp(-\frac{j2\pi kn}{N}\right|^2, 0 \leqslant k \leqslant N-1. \tag{2.31}$$

where $P[k]$ represents the PSD, T is the sampling period and $X[k]$ is the Discrete Fourier Transform of the recorded signal $x[n]$. Chen et al. (2007) also demonstrated that the magnitude of the peak in the PSD can be related to leak flow rate, shown through high Pearson's Correlation Coefficients between the feature and leak flow rate.

As leak VAE signals are non-stationary (Ahadi and Bakhtiar, 2010), approaches with good representation of non-stationary signals in both time and frequency domains could provide informative features. Empirical Mode Decomposition (EMD) is an adaptive time-frequency technique and has been shown to provide good estimates of time-frequency signals (Ghazali et al., 2010), and therefore may provide the required time-frequency resolution for leak signal feature extraction. EMD breaks down a signal into different Intrinsic Mode Functions (IMFs), which represent a different frequency

content of the signal. The EMD method as a process of sifting is described below:

1. Find extrema of signal $x'(t)$.
2. Find lower and upper signal envelopes connecting the minima (cf. minima) and maxima (cf. maxima), $e_min(t)$ $(cf.(e_max(t))$ by interpolating (using spline interpolation).
3. Calculate the signal mean between upper and lower envelopes, $m(t) = (e_min(t)+e_max(t))/2$.
4. Subtract the calculated mean to obtain 'modulated oscillation'(Mandic et al., 2013), $d(t) = x'(t) - m(t)$.
5. Apply stopping criteria (Mandic et al., 2013). If $d(t)$ satisfies stopping criteria, let $d(t)$ become IMF_m. Revert back to step 1 and subtract the new IMF from the original signal $(x'(t))$, so $x'(t) := x'(t) - IMF_m$.
6. Continue sifting until IMF calculated in step 5 becomes a monotonic function.

EMD has been used to successfully extract signals in noisy environments (Huang and Shen, 2005). As leak signals on plastic pipe generally have a low signal to noise ratio (Humphrey et al., 2012) EMD is likely to be particularly useful. EMD-based decomposition was used by Sun et al. (2016) in combination with the signal RMS to classify the aperture of leaks in gas pipes, finding that the RMS of individual IMFs, produced from EMD provided good separation between circular holes in gas pipes of different diameters.

However, EMD alone suffers from a mode mixing problem due to IMF rectification and a signal may be separated into the same IMFs (Wu and Huang, 2005) resulting in inaccurate measurements. Huang et al. (1998) suggest mode mixing could lead to the loss of signal meaning as well as aliasing in time-frequency domains. This problem can be resolved by overcoming signal intermittency (Wu and Huang, 2005), leading to the development of the Ensemble Empirical Mode Decomposition (EEMD) by Wu and Huang (2005). Essentially, EEMD decomposes the signal via the EMD method, but during each decomposition process Gaussian white noise of finite amplitude is added to the signal:

$$Ensemble : \{S_n(t)\}_{n=1}^{N} = x(t)\{w_n(t)\}_{n=1}^{N} \tag{2.32}$$

where $\{w_n(t)\}_{n=1}^{N} N(0, \sigma)$ is the generation of Gaussian white noise and $x(t)$ is the recorded leak signal. As the additive noise is different for all decomposition levels, no mixing occurs (Zhang et al., 2010). Due to the higher accuracy in generating representative IMFs, EEMD is a preferred option for feature extraction over EMD. A similar method was proposed by Si et al. (2016) who used EEMD to decompose signals into IMFs with different energy for classification with LS-SVM.

The concept of Shannon entropy was introduced in order to characterise system complexity where more random, discorded systems have higher information entropy. Sheng et al. (2016) used local mean decomposition followed by Shannon entropy and SVM in order to classify bearing running state. Sun et al. (2016) also used the Shannon Entropy of individual IMFs following EMD to recognise leak apertures. Therefore, Shannon Entropy may be a useful feature in describing leak flow rate, leak area and leak shape.

2.3.7 Feature redundancy methods

The techniques used in order to classify/predict leak signals can involve the use of a number of features. However, redundant features can increase computational cost and there is a possibility of overfitting, therefore redundant features need to be removed. Techniques known as feature selection, or feature redundancy methods can often handle a large number of features and select those that are most relevant for predicting a response. This is an important step in machine learning, where the removal of redundant features reduces computational cost and increases model accuracy. As this study is the first of its kind, it is not known which features will have the greatest use to the model and therefore feature selection becomes more important.

A variety of methods exist for subset selection and have been reviewed by numerous authors (Langley, 1994; Miller, 1990). Brute force methods such as exhaustive search involve assessing all input combinations and then identifying the subset which provides the greatest accuracy, but can have high computational cost and there is a possibility of overfitting. So called "greedy methods" include forward selection and is a more conventional feature selection algorithm. Forward feature selection methods involve measuring validation error and the best individual feature is identified. The best subset of two components is then found and continues finding the best combination of features. The greedy forward search algorithm involves a finite difference calculation, and details are given in Guyon (2003). This method works by beginning with one feature, and additionally adding all the other features incrementally. The model is evaluated for accuracy following the addition of another feature, and the feature is only kept if it increases model accuracy (Shardlow, 2002). However, a limitation of this method is that, as it is a greedy solution, it may not find the optimal set of features but does locate a set of features which are useful, as well as looking at the combination of features as each feature is not assumed to be independent of each other (Shardlow, 2002).

2.4 Summary of literature review

The literature presented herein has reviewed the current state of the art in leak detection using acoustic and vibration methods. A number of research gaps have been identified:

1. The generation of the leak noise is not fully established due to poor experimental methodologies in the literature. Therefore, it is currently not established whether the noise is generated from turbulence around the leak, cavitation or a combination of both.

2. The fundamental factors that influence a leak signal and therefore the factors that may affect the ability to detect the leak are not understood due to limitations within experimental methodologies presented within the existing literature. Moreover, the majority of controlled experiments analyse noise signatures of leaks that are not common in plastic pipes in real WDS, and are therefore not representative of real leaks. However, the literature review was able to identify some critical factors that are expected to influence leak signals, and these are identified as: (1) leak flow rate; (2) leak area; (3) leak shape; (4) surrounding backfill; and (5) pipe material.

3. No method is currently available in the literature which uses the VAE signal to predict leak flow rate, leak area and leak shape in WDS, despite the potential cost savings due to reduced non-revenue water. The reasons for the lack of method may be due to the fact the a number

of factors influence a leak signal meaning this is a very complex problem. Moreover, it is difficult to generate real leaks within a laboratory and obtaining reproducible leak signals in real WDS is a very difficult problem due to the complex nature of real WDS.

4. Due to the possible benefits of leak flow rate, leak area and leak shape estimation, a number of alternate models are reviewed which could be utilised.

Chapter 3

Methodology

3.1 Introduction

Following an exhaustive review of the literature, a number of research gaps have been highlighted and summarised at the end of Chapter 2. This enabled the development of the Aims and Objectives, with the overall aim of developing a tool to predict leak flow rate, leak area and leak shape from the VAE signal. The purpose of this chapter is to describe the methodological approach of the work presented in this research in order to achieve the Aim and Objectives described.

As discussed in Chapter 2, the characteristics of a leak signal have been shown by other authors to be governed by a number of variables. This makes the prediction of a leak's flow rate, area and shape from the VAE signal difficult due to the influence of these wide ranging variables on leak signals, and the way that they interact with each other. Some of the variables which influence the ability to predict leak flow rate, leak area and leak shape are similarly factors which also influence the accuracy of leak noise correlation. Of those reviewed in the literature, it was found that leak flow rate, area, shape, backfill and pipe material were noted as important variables influencing the leak signal. However, the degree to which these variables influence leak signals is not well understood as they have not yet been fully investigated in a controlled and systematic way. This is largely due to limitations with the experimental methodologies presented in those studies, as the effects of each of the variables on the leak signal have not been fully isolated. Moreover, there is a lack of studies based on real WDS. Therefore, the experiments carried out in this research will be designed in such a way that it is possible to isolate the individual variables that are thought to influence the characteristics of leak signals, creating a vast and unique data set. The work presented in this chapter therefore consists of 2 main parts:

1. Laboratory experiments in order to quantify how leak flow rate, leak area, leak shape and backfill type influence leak signals.
2. Field work experiments, measured in real complex WDS, in order to quantify the effect of pipe material on leak signals.

For clarity, all leak signal measurements are undertaken using accelerometers unless otherwise stated.

This chapter also presents the initial design of the pipe rig and also demonstrates a number of experiments which were undertaken in order to evaluate the status of the pipe rig during the the commissioning phase.

3.2 Laboratory experiments

3.2.1 Pipe rig description

A specific, uniquely designed pipe test rig was required in order to isolate the variables thought to influence the characteristics of leak signals, and to achieve the Aims and objectives of this study. Therefore, a novel, state-of-the-art experimental pipe rig was designed and termed the 'LiVE' pipe rig (Leaks in ViscoElastics), which is located at the University of Sheffield, UK. The pipe rig was required to be designed in such a way to minimise the possibility of signal reflections from other sources (e.g. the number of valves, fittings, sharp bends etc. was limited), in order to ensure the experimental methodology was rigorous. This also provided assurance that any observed variations in signal characteristics were actually due to the designed experimental methodologies and not because of physical features of the pipe rig (e.g. due to the presence of bends and fittings which can influence leak signals (Pal, 2008). The pipe rig was also required to fit within the limited available space whilst also being flexible in design, allowing for the change of leak shapes, sizes and backfill types.

Figure 3.1 shows a plan view schematic of the designed pipe rig used for all of the laboratory based experiments. The rig consists of approximately 26 m long pipe loop, with a 63 mm outer diameter 12 bar rated Medium Density Polyethylene (MDPE) pipe. MDPE pipe was chosen as plastic pipe is becoming more regularly installed by UK water companies (Fox, 2016) and is a significant challenge in leak detection (Pal, 2008). Water is supplied from an upstream reservoir (0.95 m^3 volume) (part 'A', Figure 3.1) by a 3.5 kW (Wilo, Burton Upon Trent, U.K.) MVIE variable speed pump set at 15000 revs per minute (part 'B', Figure 3.1). Water then passes through a butterfly valve (part 'C', Figure 3.1). and through a magnetic flow meter (Flow Systems 91DE) measuring system flow rate upstream of the leak (part 'D', Figure 3.1).

A removable *'test section'* is located in the middle of the pipe rig, which consists of a 5.5 m long section of MDPE pipe. The test section allows for the alteration of leak area and leak shape by removing and replacing the test section at flange plates located at either end of the 5.5 m test section (locations *H* and *F* in Figure 3.1). A new test section piece is installed every time a new leak area or leak shape is investigated. The test section passes through a rectangular box (part 'G', Figure 3.1) measuring 0.5 x 0.5 x 0.5 m, which is shown in Figure 3.2. The inclusion of the rectangular box in the middle of the test section allowed for the variation of backfill types, where alternate backfill types can be inserted in to the test section box. An overflow was positioned 10 cm from the top of the test section box in order to keep the level of saturation the same for each test. The wavespeed in the pipe rig is estimated to be 347 m/s using the theoretical calculations given by Almeida et al. (2015a).

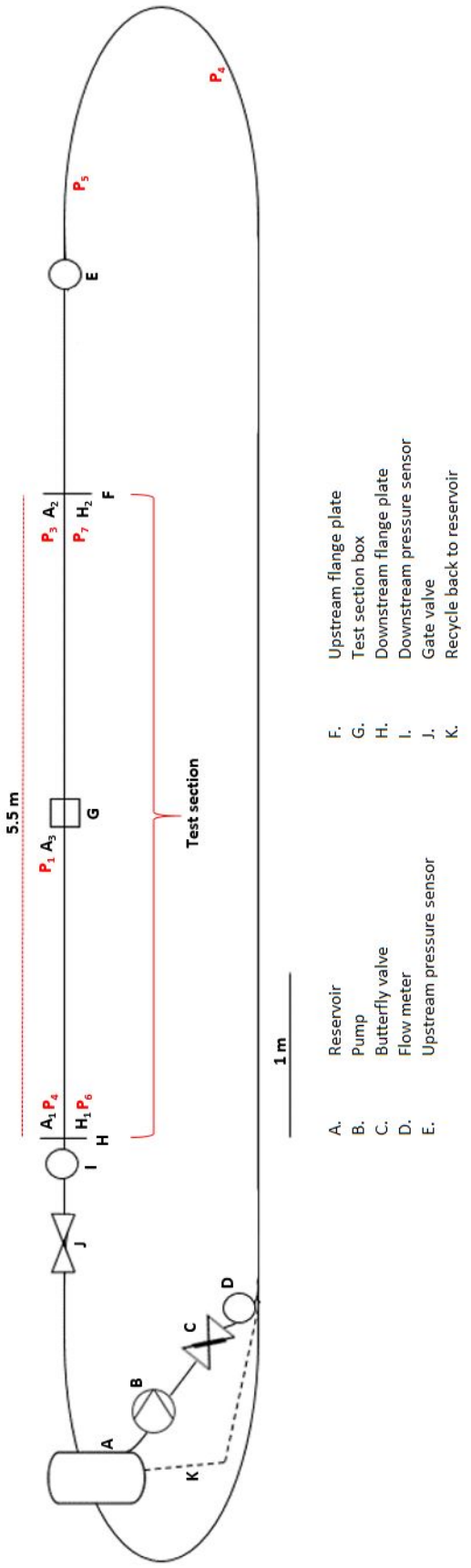


Figure 3.1: *Schematic of the LiVE pipe rig used for the laboratory experiments throughout this study.*

Figure 3.2: *Photograph of the test section box located within the LiVE pipe test rig.*

3.2.2 Pipe rig signal processing and data acquisition

A wide variety of sensors are available for leak detection, including hydrophones, geophones, accelerometers, GPR etc. Both accelerometers and hydrophones are commonly used for leak detection via VAE methods (Gao et al., 2005) and therefore both of these sensors were utilised in this study to capture leak signals. Two hydrophones (Bruel and Kjaer type 8103, 50 x 9.5 mm, sensitivity 0.162 pC/Pa) were placed either side of the leak, approximately 2.25 m and 2.41 m from the source of the leak (Positions H_1 and H_2 in Figure 3.1). Three accelerometers were also used (PCB Piezotronics 393B12 Sensitivity 10 V/g), and the position of these was varied depending on aim of individual experiments. For most experiments, one accelerometer was placed approximately 30 cm away from the leak, and another two accelerometers were placed either side of the leak, were accelerometer 2 and accelerometer 1 were placed at distances of 2.25 m and 2.41 m from the leak respectively. The positions of the accelerometers are detailed in Figure 3.1, and accelerometer 3, accelerometer 2 and accelerometer 1 are shown by A_3, A_2 and A_1 respectively, and placed in positions P_1, P_3 and P_2 respectively. Therefore, the accelerometers were placed at the same distance as the hydrophones for some tests. The accelerometers were occasionally moved to different positions along the pipe rig, for the purpose of experimental tests investigating a specific purpose. For example, when investigating the influence of the pipe transmission on leak signal attenuation, accelerometer 2 was moved to position P_4 (8 m from the leak). A summary of the sensor placement positions and their respective distances from the leak for each experiment undertaken is given in Table 3.1. The accelerometers were placed on a specially design mounting (Figure 3.3a and 3.3b), so that the mounting would come in direct contact with the pipe wall (i.e. the mounting was stuck to the pipe wall with mounting wax). The hydrophones were inserted directly into the pipe rig with a plumbed fitting, shown in Figure 3.3c.

Accelerometers and hydrophones were both set to record signals at a sample rate of 2500 Hz in accordance with the Nyquist theorem (Landau, 1967). Leak signals were recorded for 5 seconds and 20 repeat samples were taken for each test. All sensors were powered by a current source unit (Dytran Instruments type 4102C). Signals were then passed through a 6 m integral cable to a 2 channel CCLD conditioning amplifier. A data acquisition board (National Instruments Compact DAQ cDAQ-9714) digitised signals, which were then downloaded to a desktop computer (Lenovo, Windows 7). Following data acquistion, signals were then imported and processed in Matlab. The signal was multiplied by a Hanning window with 50% overlap (Almeida, 2013), which zero-values the product outside of the window (Pal, 2008). A 4^{th} Order Butterworth bandpass filter at set points <10 Hz and >1000 Hz was used to filter the recorded signals. Traditionally, leak signals on plastic pipe are reported as low frequency (<200 Hz (Brennan et al., 2006a; Almeida et al., 2014a;

Table 3.1: *Position of sensor positions relative to leak.*

Sensor	Position	Distance from leak	Test	Sampling frequency	Comment
Accelerometer No.3	P_1	30 cm	Characterising leak signal at source and effect of pump speed	2500 Hz	Fixed in position
Accelerometer No.2	P_2	2.25 m	Transmission further away from leak	2500 Hz	Position varied
Accelerometer No.1	P_3	2.41 m	Transmission further away from leak	2500 Hz	Fixed in position
Accelerometer No.2	P_4	8 m	Signal attenuation investigation	2500 Hz	Position varied
Accelerometer No.2	P_5	5 m	Effect of bends and fitting	2500 Hz	Position varied
Hydrophone No.1	P_6	2.25 m	Measurements using a hydrophone	2500 Hz	Fixed
Hydrophone No.2	P_7	2.41 m	Measurements using a hydrophone	2500 Hz	Fixed

Gao et al., 2002; Pal et al., 2010)). However, the frequency content of the leak is likely a function of the distance the sensor is placed from the leak as the pipe acts as a low pass filter (Almeida et al., 2014a). Therefore, when measuring next to the leak, at such close proximity, the frequency content is generally unknown and therefore a higher bandwidth filter is likely required, and thus frequency content >1000 Hz was filtered out of the signal. Additionally, two pressure sensors (Gems Plainville 2200) were placed at approximately 3 m upstream and 3 m downstream of the leak, measure system pressure at a sampling rate of 2000 Hz (parts 'E' and 'I' in Figure 3.1).

(a) ***(b)*** ***(c)***

Figure 3.3: *Photographs of sensors: (a) photograph of one of the accelerometers used, with mounting piece and ruler for scale; (b) accelerometer attached to specially designed mounting enabling the sensor to sit on the pipe wall; and (c) hydrophone inserted in to pipe.*

3.2.3 Commissioning of the pipe rig

As the pipe rig is a new installation and specifically designed for this study, a commissioning phase was required in order to better understand the pipe rig and untangle any problems that the future study may face. It is also important to understand the level of ambient noise and noise produced from the pipe rig itself (e.g. noise from the pump, flow related noise etc.) before any experiments take place in order to establish a reference condition. For clarity, background noise in this thesis is characterised as signals that are present within signals recorded with no leak in place. The following tests encompass a set of experiments which are designed to identify any issues that may effect later experimental results or lead to conclusions drawn from results skewed by the presence of features occurring due to the pipe rig.

3.2.3.1 A: Effect of system pressure

Initially, accelerometers were placed in position P_1 (30 cm from where the leak would normally be drilled in to the pipe test section, Table 3.1 and Figure 3.1) with the pump running at 15000 rpm. In this case, the VAE signals of the pipe rig is recorded and analysed with no leak in place. Recording the system noise with no leak in place at various levels of system pressure creates a reference data set containing only background noise at different system pressures. Figure 3.4 demonstrates a frequency domain accelerometer signal of the pipe at 5, 8 and 11 m head with no leak in place, solely recording background noise. In the case of 5 and 8 m head, there appears to be negligible differences in frequency spectra, suggesting that system pressure has little effect on received accelerometer signals at these set system pressures.

At 11 m head, a peak appears at approximately 318 to 416 Hz, which opposes the trend for the other system pressure set points observed in Figure 3.4. It is possible that the modifications to signal spectra within this frequency band is related to valve screech as the downstream valve was closed to increase system pressure, where the part-closed valve could have lead to a noise as the water slipped through the valve. An assessment of whether or not this was valve screech due to the part-closed valve was carried out by beginning with the downstream valve fully open (zero clockwise turns, no valve screech) and then incrementally closing the valve 1 whole turn clockwise (360^0) and repeating tests. The frequency spectrum results of this investigation are shown in Figure 3.5. It was found 16.5 clockwise turns are required to fully close the valve from open, resulting in zero system flow. Figure 3.5 shows that the erratic noise observed between 318 to 416 Hz identified in Figure 3.4 was found to occur between 12 and 13 valve turns during this investigation. However, this peak seems to settle down back to the 'normal' trend when turning the valve between 14 and 15 whole valve turns. This suggests that between 12 and 13 clockwise valve turns there appears to be valve screech which is later removed by either opening or closing the valve further. These results demonstrate that the screeching valve at these system pressure will influence any recorded signals, however it is avoidable by neglecting samples at 12 and 13 valve turns.

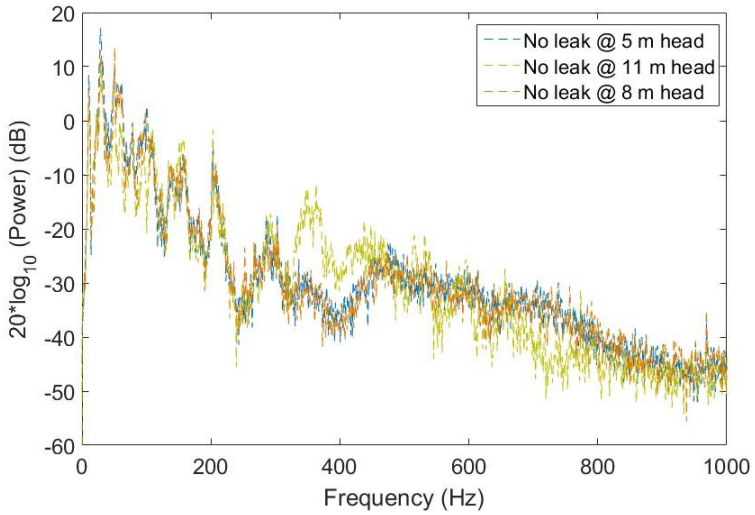


Figure 3.4: *Changes in frequency and amplitude at different system pressure with no leak in place. Accelerometer is positioned 30 cm from the leak.*

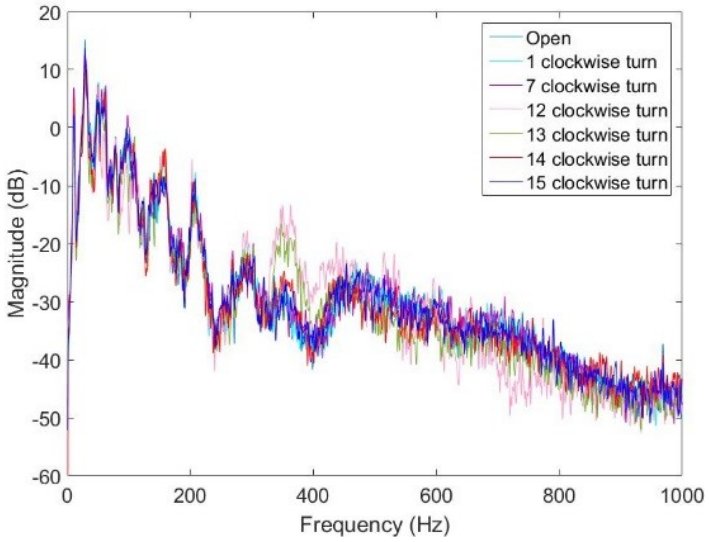
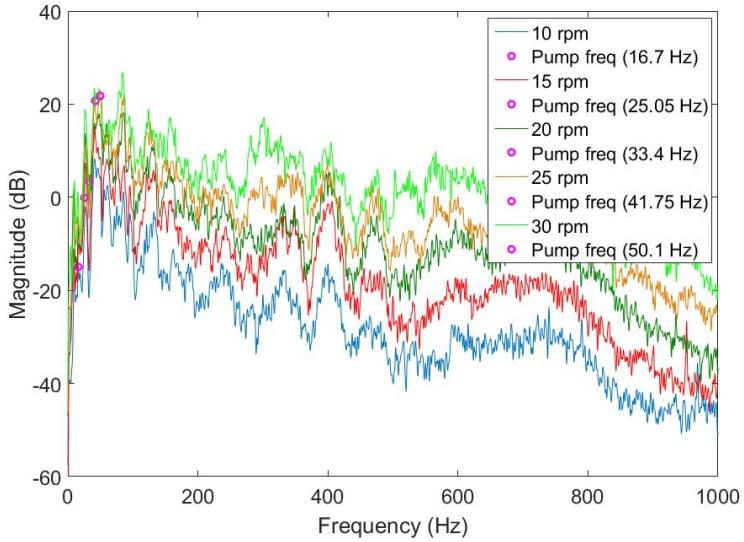


Figure 3.5: *Changes in frequency and amplitude at different system pressure with no leak in place, referenced by the number of 360^0 valve turns. Accelerometer is positioned 30 cm from the leak.*

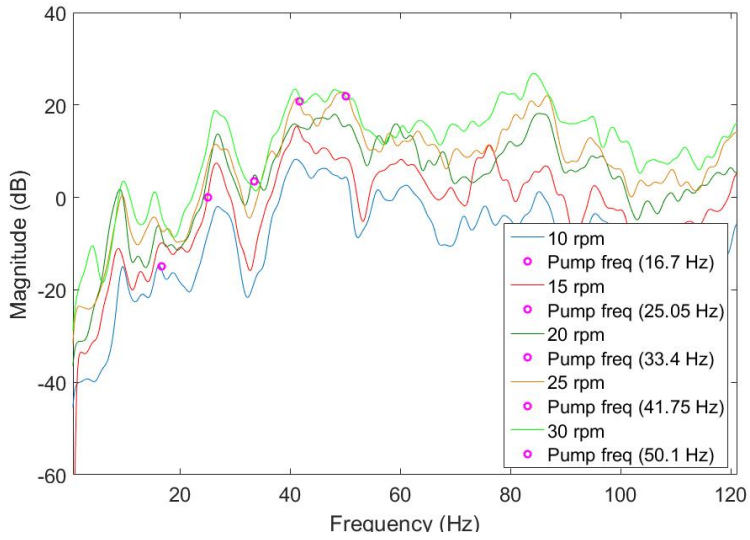
3.2.3.2 B: Effect of pump speed

It is known that pump activity can cause a level of system noise which can interfere or 'mask' the leak signal (Murvay and Silea, 2012). Moreover, both Figure 3.4 and Figure 3.5 appear to display some distinct cycling, and as there is no leak in place, this may be due to the centrifugal pump used to circulate water around the pipe rig. Therefore, the effect of the pump on the recorded signals needs to be understood during the commissioning phase. The investigation in to pump noise was carried out by deploying Accelerometer 3 in to position P_1 (Table 3.1) and increasing the pump speed to effectively increase noise from the pump. The pump speed was varied between 10000 rpm and 15000 rpm to capture a range of noise from the pump. The accelerometer readings of the pump at the various speeds are expressed in the frequency domain in Figure 3.6a. Figure 3.6b shows the same plot as Figure 3.6a but with the axes scales adjusted in order to

better visualise the frequency the pump operates at (a 'zoomed in' version providing a greater degree of granularity and detail at the lower frequency bands where the pump operates). The conversion of pump motor rpm to operating frequency (Hz) was calculated by $\frac{\mathrm{Pump}_{\mathrm{rpm}}}{60}$. The operating frequency of the pump at the various pump speeds is represented by pink dots in both Figure 3.6a and Figure 3.6b. The calculation of pump operating frequency is useful as it allows for the identification of peaks in signal spectra which are due to the pump (and therefore not due to the leak). Analysis of Figure 3.6a and Figure 3.6b shows that evidently, increasing pump speed increases the level of ambient noise. Therefore increasing pump speeds is not suitable for any future experiments as accelerometer spectrum's will not be comparable when comparing results with different pump speeds. Therefore, the pump should be kept consistent throughout the experiments.



(a)



(b)

Figure 3.6: *(a) Spectrum of signals with changing pump speed; (b) is the same plot as (a) but axes scales are changed. Pink dots represent the pump motor operating frequency in Hz. Accelerometer is positioned 30 cm from the leak in both figures.*

3.2.3.3 C: Effect of pipe bends and fittings

Whilst the pipe rig was designed in order to minimise any situations where bends, valves and fittings that may interfere with the leak signal, space limitations within the hydraulics laboratory required the pipe to be curved in an oval shape. It was reported by Pal (2008) that the presence of bends and fittings can cause changes to the signal spectra. These differences are likely to be generated by fluid turbulence around a bend or fitting which in turn creates a vibration through the fluid and the pipe wall. Therefore, investigation in to the effect of these features is required as it is important to fully understand which components of the system may generate a difference in signal spectra due to noise created by bends and fittings.

The influence of bends and fittings was investigated by moving the accelerometers into several positions along the pipe rig, in to positions P_1, P_2 and P_5 (Table 3.1) which were strategically selected locations as there may be an influence of bend and fittings at these positions. Position P_1 has the accelerometer positioned 30 cm from the leak, P_2 accelerometer is placed 2.25 m away from the leak behind the flange plate (connecting the pipe rig to the test section) and P_5 accelerometer placed 5 m from the leak just after the pipe bend. Therefore, any differences between P_1 and P_2 should highlight the effect of the flange plate and differences between P_3 and P_1 will highlight turbulence around the bend. The results from this investigation are shown in Figure 3.7. It appears that there are some slight differences in spectra due to the presence of bends and fittings, however, these appear to be negligible differences and do not appear to be of any significant concern. Therefore the design of the pipe rig has minimised any noise due to the presence of fittings. Note that this test is carried out with no leak in place and therefore does not investigate how the presence of bends or fittings may effect the transmission of the leak signal, but does establish the effect of the rigs physical parameters on the recorded vibration signals.

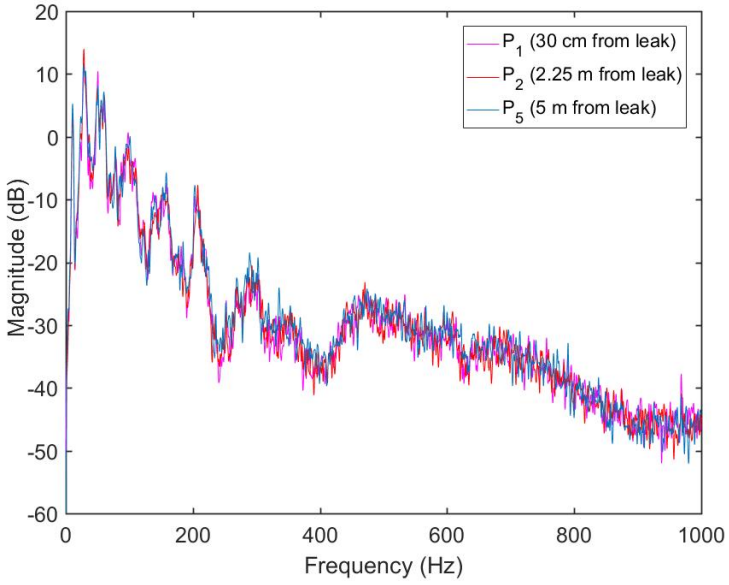
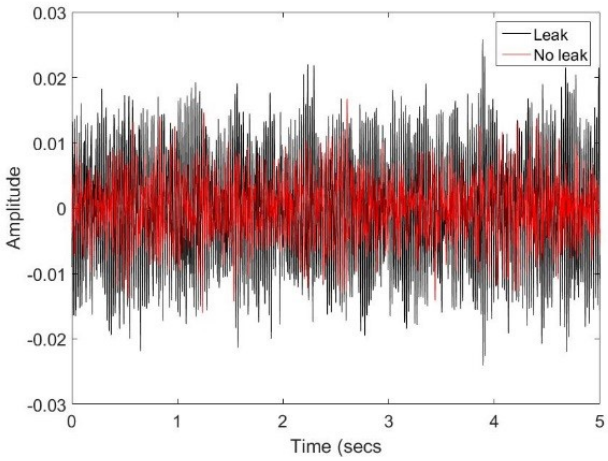


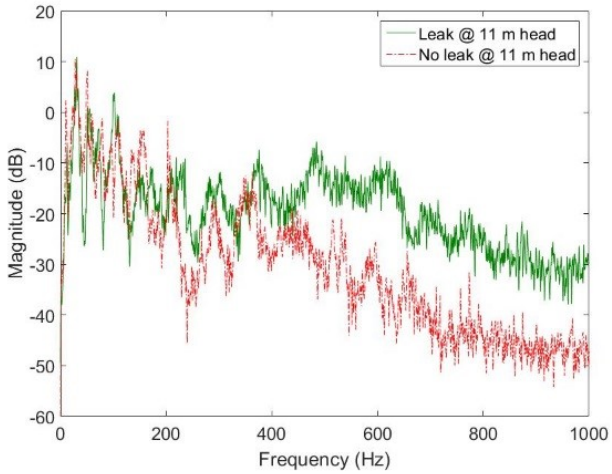
Figure 3.7: *Influence of pipe bends, measurements taken at various points around the pipe rig. Accelerometer position is varied and is given in the figure legend.*

3.2.3.4 D: Introduction of a leak as a noise source

As the system background noise and the physical effects of the pipe rig have now been investigated without a leak in place, a leak can be added in to the system in order to establish the effect of this on signal spectra compared to the background noise case. A leak was introduced by replacing the no-leak test section with a new test section which included a 3.5 mm round hole, created with a 3.5 mm drill bit. Figure 3.8a and Figure 3.8b demonstrate the time domain and the frequency domain spectrum of the leak at 11 m head respectively. The 'no leak' measurement has also been included for comparison. The introduction of a leak has caused water to discharge through the hole. In turn, this has created additional vibration on the pipe wall and in the fluid, which is visible in Figure 3.8a and Figure 3.8b by comparing the differences in signal spectra between the signals representing no leak and with a leak. The time domain signal comparison shown in Figure 3.8a clearly shows a distinction between the no leak and leak signals with greater amplitude signal generated through the addition of a leak. The frequency domain comparison shown in Figure 3.8b shows that the leak and no leak patterns appear remarkably similar at frequencies $<$214 Hz. However, on closer examination, the difference between leak signal and no leak signal is more apparent at approximately 28 Hz. Therefore, any signal $<$28 Hz can be considered background noise as the addition of the leak does not appear at frequencies $<$28 Hz.



(a)

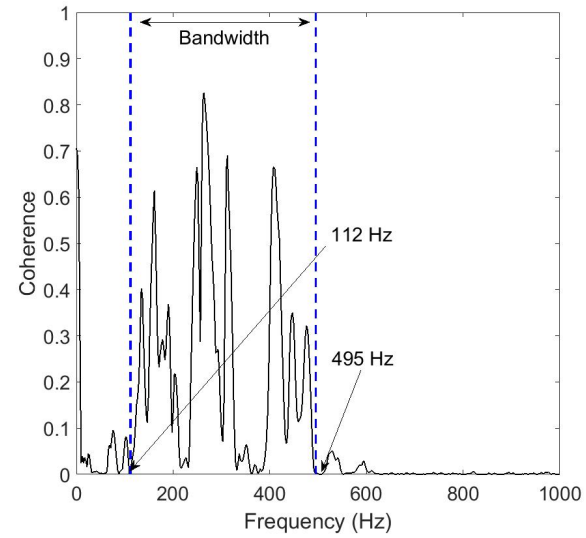


(b)

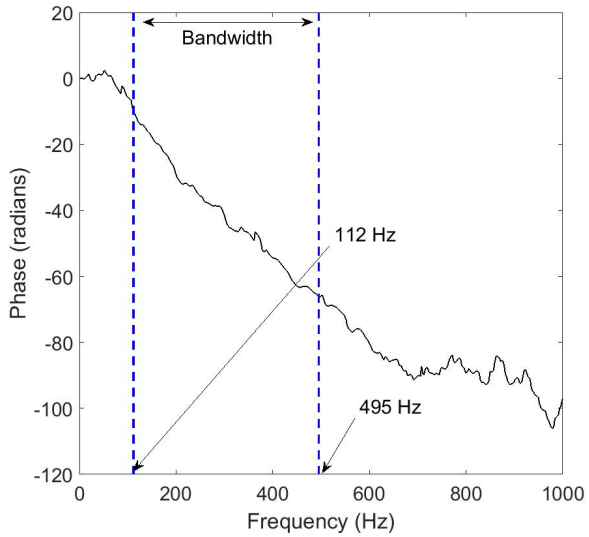
Figure 3.8: *Leak and no leak comparison: (a) time domain; and (b) frequency domain representations. Accelerometer is positioned 30 cm from the leak.*

The intitial assessment of leak signals from the laboratory also took place by observing both the coherence, phase and cross correlation of a leak signal, which are common methods of analysing leak signals shown by a number of authors, for example Pal (2008), Almeida et al. (2014b) and Gao et al. (2002). Two accelerometers were placed directly on the pipe at positions away from the leak: Accelerometer 1 at position P_2, 2.25 m away from the leak; and accelerometer 3 at position P_3 2.41 m from the leak (Table 3.1). The coherence and phase plots of both accelerometer signals is plotted in Figure 3.9. The coherence plot in Figure 3.9a shows that coherence was more pronounced between a specific bandwidth, between 113 and 495 Hz (note that sensors here are not placed next to the leak and therefore there is an effect of transmission on the leak signals). The peak in coherence is at 0.83 and occurs at 293 Hz. The coherence estimate appears to oscillate between a high level of coherence and zero coherence within the same bandwidth (i.e. there is not a high band of coherence within one particular bandwidth as seen by other authors (e.g. Almeida et al. (2014a)). The phase spectrum shown in Figure 3.9b shows a largely linear decrease in phase with increasing frequency within the same bandwidth of higher coherence (112-495 Hz). However, within this band a number of distortions (or 'ripples') in the phase spectra are still present, and it is not entirely linear. These distortions in the phase spectra appear at similar frequencies that additionally have low coherence shown in Figure 3.9a. Therefore, when recording leak signals at distances away

from the leak an effective method of pre-filtering signals will be to apply a bandpass filter which maximises frequency bands with high coherence and linear phase with frequency.



(a)



(b)

Figure 3.9: *Assessment of leak signals in the laboratory using accelerometers: (a) Coherence estimate; and (b) Phase estimate. Accelerometer 2 at position* P_2*, 2.25 m away from the leak and accelerometer 1 at position* P_3 *2.41 m from the leak.*

3.2.3.5 E: Cross correlation of leak signals

The cross correlation of leak signals is one of the primary methods of locating a leaks position in a pipeline, utilising VAE signals. However, the efficacy of cross correlation measurements (and therefore the accuracy of leak detection) is often influenced by a number of variables, such as other noise sources (Pal, 2008) causing reflections and distortions in the signal (Brennan et al., 2006a). It is of great importance in this study to investigate whether the leak signal can actually be correlated and whether the leak's location within the pipe rig can be localised accurately, especially investigating whether any external noise sources interfere with the cross correlation.

As mentioned in Chapter 2, Gao et al. (2006) compared a number of GCC estimators with the BCC to enhance the cross correlation of leak signals. However, there still remains no consensus over which GCC methods provide the most accurate results. In order to assess the best GCC estimator for this test rig, Accelerometer 2 was placed in position P_2, 2.25 m away from the leak and accelerometer 1 was placed in position P_3, 2.41 m from the leak. The BCC is compared with the PHAT, SCOT and ML estimators (these algorithms are discussed in Chapter 2). This involves using the same recorded leak signal, but performing some additional signal processing which varies depending on the estimator selected. Due to the fact that the data was normalised (and therefore the peak assumed to be the leak is maximised), the BCC contains a strong peak (Figure 3.10a). However, the normalised peak denoting the maximum cross correlation coefficient observed is surrounded by a number of other peaks which could be due to the leak, noise or reflections. These additional peaks in close proximity to the peak due to the leak introduce a level of ambiguity, and therefore there is a reduced level of confidence in that the main peak may actually be due to the leak. In the case of the other estimators, the PHAT (Figure 3.10b), SCOT (Figure 3.10c) and ML (Figure 3.10d) managed to attenuate the surrounding peaks and therefore distinguish the peak due to the leak better. This shows that the GCC estimators would provide more reliable cross correlations between noisy leak signals. In the case of the of the PHAT estimator, it appears to add some extra noise in to the cross correlation. This study has shown that the leak can be localised to the correct position and that the SCOT estimator is likely to yield the most accurate estimates of leak position within this test rig.

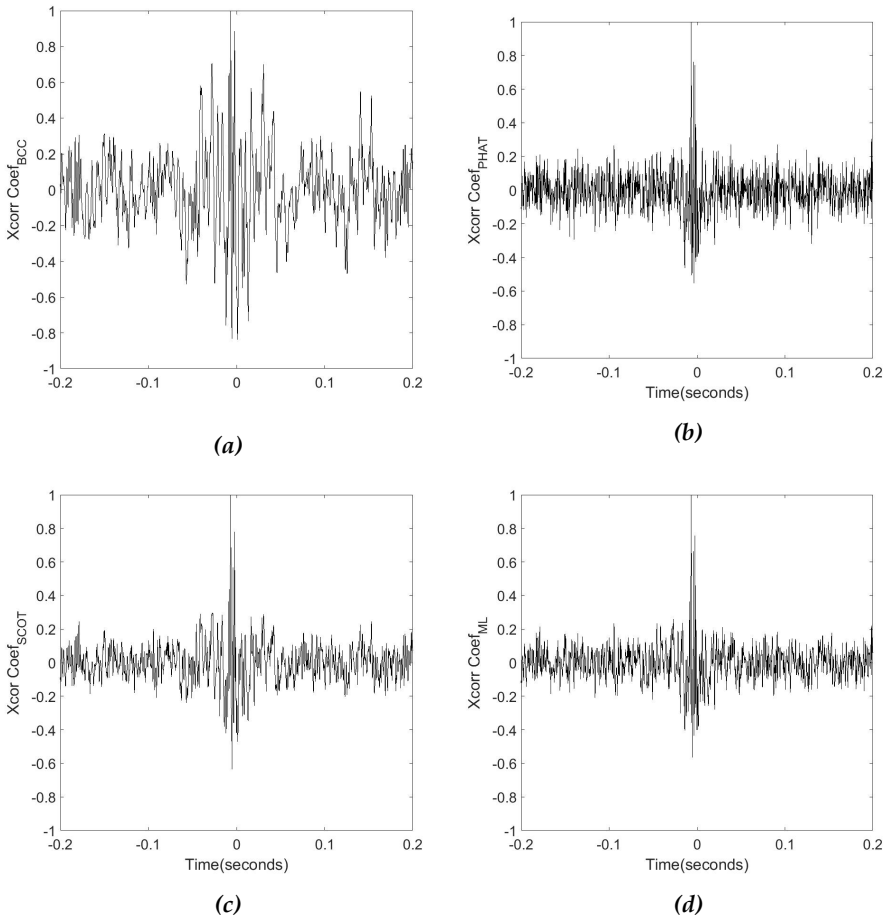


Figure 3.10: *Comparisons of weighting estimators used for the cross correlation of a leak signal: (a) BCC; (b) PHAT; (c) SCOT; and (d) ML estimators. The positions of the two sensors used to correlate the signals are: accelerometer 1 in position* P_2*, 2.25 m away from the leak and accelerometer 3 in position* P_3 *2.41 m from the leak.*

3.2.3.6 F: Attenuation of leak signals

In real pipelines, propagating VAE signals are dampened due to the transmission of energy into surrounding media, dissipative processes where acoustic energy is converted to heat energy. Therefore, damping can have a significant effect on the attenuation of signals. As sensors are moved further away from the leak, there will be a loss of leak signal energy making it harder to estimate the leak's location accurately, which is particularly true on plastic pipe. Moreover, distance from the leak source is likely to have an impact on the characteristics of the leak signal. Transmission coefficients can be applied in order to demonstrate how the pipe will attenuate leak signal, given by Almeida (2013), where Equation 2.9 (discussed in Chapter 2) is shown to calculate the attenuation rate of leak signals. Some parameters are necessary for the calculation of the theoretical attenuation factor which relate to the pipe-rig, and these are given for the LiVE pipe-rig used in this study in Table 3.2.

Table 3.2: *Parameters relating to the pipe-rig used for calculating theoretical pipe attenuation factor.*

Parameter	Value
Pipe thickness	5 mm
Free fluid wavespeed	1500 m/s
Young's Modulus	2 x 10^9 N/m^2
Fluid bulk modulus	2.2 x 10^9 N/m^2
Pipe Diameter	63 mm
Damping factor n	0.022
δ is therefore	1.3 x 10^{-4}

Using the attenuation factor calculated from Equation 2.9 and the parameters given in Table 3.2 it is possible to calculate a theoretical estimate of attenuation for given frequencies. The subsequent calculation of Equation 2.9 shows that higher attenuation occurs at higher frequencies and a linear increase in attenuation with frequency will be observed. From the calculation, the attenuation is estimated at 200 Hz is equal to 1.49 dB/m, therefore at 10 meters the total attenuation is 10.49 dB at this frequency. The ability of the theoretical model to predict attenuation levels can be compared with the pipe rigs actual attenuation using experimental results. The actual (measured) attenuation was recorded by taking the transfer function of the leak noise between two sensors by keeping acclerometer 3 in position P_1 (30 cm from the leak, Table 3.1) and moving acclerometer 2 to position P_4 (8 m from the leak, Table 3.1). The plotted comparison between theoretical attenuation and actual attenuation is demonstrated in Figure 3.11. A least squares regression line was drawn through the actual (recorded) attenuation data and it is shown that the attenuation of different frequencies is given by $0.0013f + 0.38$, where f denotes frequency (Hz). The measured (actual) attenuation from the transfer function was found to differ at higher frequencies compared to the predicted attenuation calculated using the theoretical equations given by Almeida (2013) and Gao et al. (2004). However, at lower frequencies (>93 Hz), the theoretical values closely resemble the observed values.

The differences between the modelled and actual attenuation rates are likely as the damping factor used to calculate δ does not take into account external factors (such as backfill), which can have an influence on the damping factor (Almeida et al., 2015a). Other factors such as discontinuities in the pipeline can increase the level of attenuation, which are not included in the theoretical calculations (Pal, 2008). Therefore, it appears that for complex WDS, the attenuation factors proposed in the literature may not sufficiently describe the attenuation of the leak signal. However, for a simple linear system the theoretical attenuation factor estimates may be appropriate at lower frequencies, although it is unlikely in a real WDS that a system would contain a straight pipe section with no reflections or discontinuities.

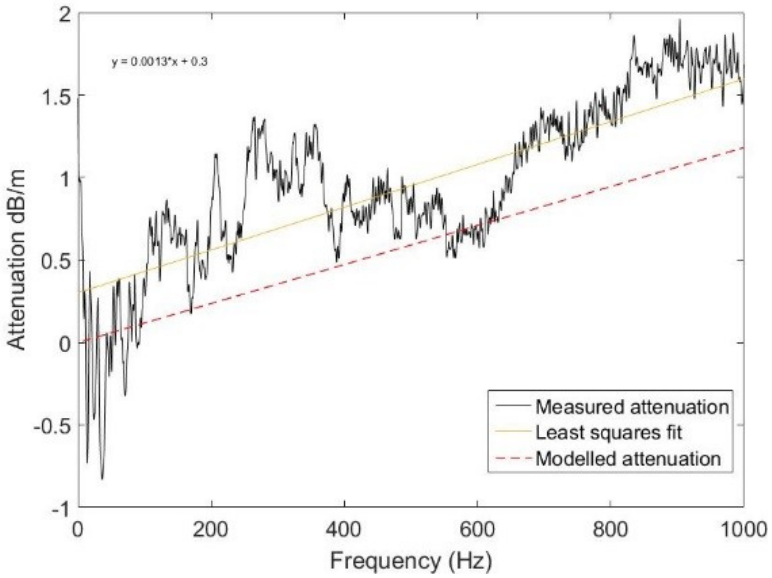


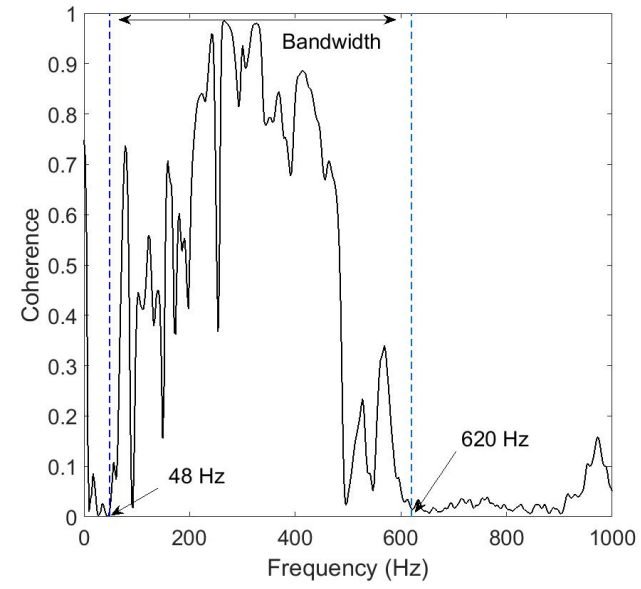
Figure 3.11: *Measured attenuation vs. modelled attenuation of the LiVE pipe rig. Accelerometer 3 position* P_1 *(30 cm from the leak) and accelerometer 2 in position* P_4 *(8 m from the leak) is used to calculate the measured (actual) attenuation of the leak signal.*

3.2.3.7 G: Comparisons between accelerometers and hydrophones

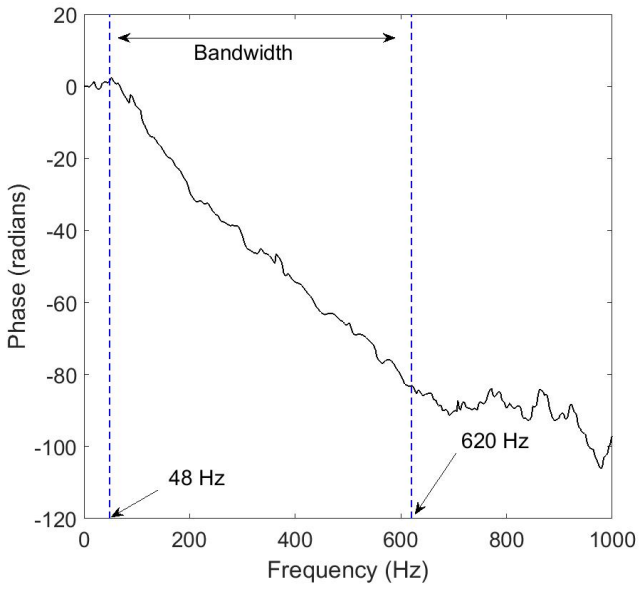
The leak spectrum was additionally measured by placing hydrophones in the same position of the accelerometers (positions P_2 and P_3). This equated to hydrophone 1 (H_1, re 3.1) being placed 2.41 m and hydrophone 2 (H_2, Figure 3.1) being placed 2.25 m from the leak. The coherence and phase plots of the leak measured with hydrophones are plotted in Figure 3.12. The plotted data is for the same condition as that plotted in Figure 3.9 as the measurements were taken simultaneously. Figure 3.12a shows that the hydrophone seemed to have a slightly larger bandwidth compared to the accelerometer signal, with strong coherence between 48 Hz and 620 Hz. Within this frequency band, the hydrophone had stronger coherence in general, especially at frequencies between 290 and 500 Hz where coherence was found to be > 0.7. Therefore, it appears that under these test conditions, the hydrophone provided better coherence estimates and a greater spectral bandwidth of leak signal. The phase estimate shown in Figure 3.12b demonstrates that the phase is also linearly decreasing with frequency between the same bandwidth of coherence (48-620 Hz). The plot also shows a number of distortions in phase as also observed in the accelerometer phase plot (Figure 3.9b). The hydrophones appeared to have much greater coherence with a larger bandwidth compared to the accelerometer signals. The fact that that the hydrophones have greater coherence agrees with studies shown by Gao et al. (2005) and Almeida et al. (2014a).

The differences noticed between the accelerometer and hydrophone are also possibly due to the experimental conditions. The ground conditions have been shown to have a strong influence on the leak signal (Muggleton and Brennan, 2004). The accelerometers are positioned on the pipe wall, and therefore are susceptible to changes in the backfill conditions. Whilst efforts were made to ensure that the ground conditions were standardised, the replacement of test sections resulted in the excavation of the pipe from the gravel backfill. Slight changes to the test conditions are possible through changes in loading, soil hydraulics and flow resistance which can have an influence on leakage dynamics (Fox et al., 2016). As the accelerometer was placed in contact with the pipe

wall, the effect of ground conditions may be more paramount and therefore may interrupt with the signal more so than the hydrophone. The distortions in the phase spectrum shown in both the hydrophone (Figure 3.12b) and accelerometer (Figure 3.9b) phase spectrum's are likely due to reflections and discontinuities in the pipe line, which was also observed by Almeida et al. (2014b). Moreover, these resonances have been shown to influence the time delay estimate (Almeida et al., 2014b) and therefore the accuracy of leak detection via noise correlation. These results are of importance as the effectiveness of the correlation between signals relies on good coherence and linear phase between sensors. However, accelerometers still remain the preferred sensor type over hydrophones in use by water companies due to the fact they are non-invasive and therefore water quality risks are negated (Pal, 2008).



(a)



(b)

Figure 3.12: *Assessment of leak signals recorded using hydrophones in Position* P_6 *(2.25 m from leak) and* P_7 *(2.41 m from leak), where (a) is a coherence estimate; and (b) is a phase estimate.*

3.2.3.8 H: Repeatability of measurements

A study by Hunaidi and Chu (1999) investigating the acoustic characteristics of leaks from a pipe test rig in Canada suggested that their results were not repeatable, however most authors studying leak signals and/or leak detection do not allude to the repeatability of their results. Some studies avoid repeat measurements as there is inherent variability due to the uncontrollable factors during leak noise correlation, such as background noise, soil temperature (e.g. (Almeida, 2013)). However, investigations of repeatability were carried out by Pal (2008), who found very repeatable results in laboratory conditions. It is of vital importance that a rigorous experimental methodology is developed which is both repeatable and reproducible.

The pipe rig developed in this study is in an isolated, temperature controlled room and factors such as background noise are limited, especially in comparison to real WDS. Therefore, it should be possible to achieve repeatable data sets if due care is taken when carrying out experiments. Elements of the experimental procedure can be controlled in order to reduce the possibility that data is not reproducible, including by the following methods:

- Sensor placement in an alternative position can result in a different signal. This can be minimised by ensuring sensors are placed in the same position each time.
- Turbulence around the leak hole is different due to the presence of swarf. This effect can be minimised by inserting and removing the drill bit from the leak hole 3 times to remove any swarf present.
- Turbulence around the leak hole changes to the adjusted angle of the water jet due to changes in backfill which is particularly apparent for gravel backfill. This case can be avoided by placing the gravel in test section box, packing it down tightly and fully saturating it with water.

During the commissioning phase, a leak's spectrum was analysed for repeatability by removing and replacing the accelerometers in to the same position and analysing the results. A similar method of analysing the repeatability of leak signals was also shown by Pal (2008). In Figure 3.13, the accelerometer has been removed and replaced 3 times with and without a leak in place in order to show how repeatable the measurements are. Due to the negligible differences in signal spectra as the accelerometer is removed and replaced, it appears that it is possible to achieve repeatable results when measuring leak signals on this pipe rig.

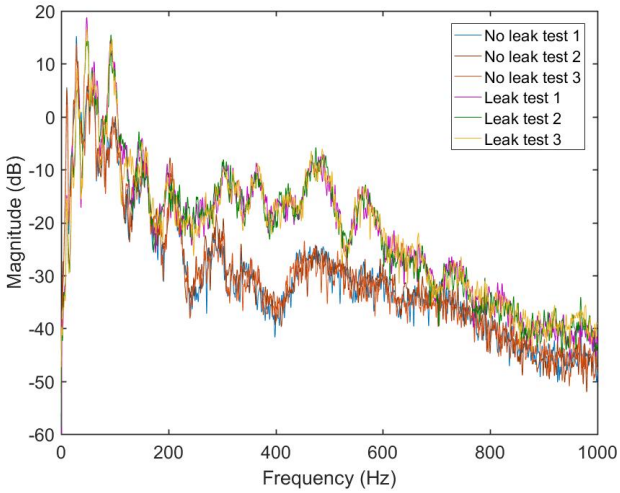


Figure 3.13: *Frequency domain plot demonstrating the repeatability measurements with and without a leak. Each additional test represents the accelerometer placed on the pipe test rig, lifted and then placed back on to the pipe in the same position. Accelerometer is placed 30 cm from the leak in position* P_1.

3.2.4 Laboratory experimental design

Chapter 2 identified that various variables influence a leak signal in different ways, and the methods by which they interact are extremely complex. However, there appears to be a lack of understanding of how these variables influence a leak signal due to the lack of control in the experiments carried out. In order to fully understand the fundamental factors influencing leak signals and to achieve the Aims and Objectives of this study, a specific set of experimental tests were required. This involved changing a number of parameters within the pipe-rig which isolated different variables, and to carry out specific experimental tests to determine the influence of these factors on the characteristics of the leak signal. Specific details of each experiment are shown in the following sections.

3.2.4.1 Laboratory preparation

Before any experimental tests took place on the pipe rig, the pump was started and left to run for 3 hours. This allowed for any viscoelastic effects of the pipe rig to equalise and the laboratory to stabilise, and this took place for all the laboratory experiments given below.

3.2.4.2 A: An investigation in to the influence of leak area and leak shape on leak signals

The influence of leak area and shape has been neglected in the literature. Commonly, round holes of varying diameters are drilled in order to increase leak area. Comparisons have been made by several authors to show this (e.g. Pal (2008)) and to demonstrate how a leak signals spectra can change with increasing hole area (Sun et al., 2016). However, as the leak area increases the leak flow rate also increases. Therefore the effect of leak area on the leak signal has not been fully isolated as the observed changes to signal spectra noted by other authors were more likely due to changes in leak flow rate rather than leak area. The same is true of leak shape, as leak shape has never been fully isolated due to changes in other variables (such as the leak flow rate varying when studying different leak shapes). Moreover, there has been no representation of real leaks in terms of VAE

signals in controlled laboratory conditions within the literature, as round holes are rare on plastic pipes in real WDS (Tayefi, 2014).

In this study, a variety of leak shapes were drilled with differing leak areas. Round holes and longitudinal slits of 4 leak areas each where drilled in to separate test sections. Additionally, two leaky electrofusion joints were created by creating an artificial void in one pipe before welding in order to create a leak more representative of "real leaks" (leaky electrofusion joints are one of the most common on plastic pipes (Tayefi, 2014)). A schematic of the void designed to create the leaky electrofusion joints is given in Figure 3.14. A drill bit was used to drill the correct sized leaks, which was inserted and removed 3 times in order to reduce the presence of swarf. Photographs of the discharge patterns of these leak shapes are presented in Figure 3.15. As there is a resultant change in leak flow rate (and therefore change in leak signals) as leak area increases, the system pressure was adjusted in order to maintain a level of control over the leak flow rate exiting through the leak. The system pressure was varied by adjusting the downstream gate valve, which also allowed for leak area and leak shape to be individually isolated. When the shape, backfill and leak flow rates are kept consistent, the only changing parameter is leak area. In cases where the leak flow rate, area and backfill are kept consistent, the only varying parameter is leak shape. Details of the leak shapes and leak areas studied for different leak types are provided as part of a data summary in Table 3.3 later in this chapter.

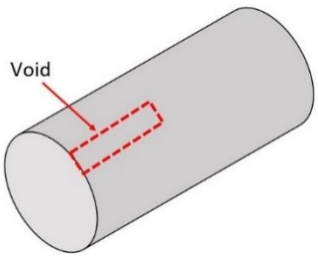
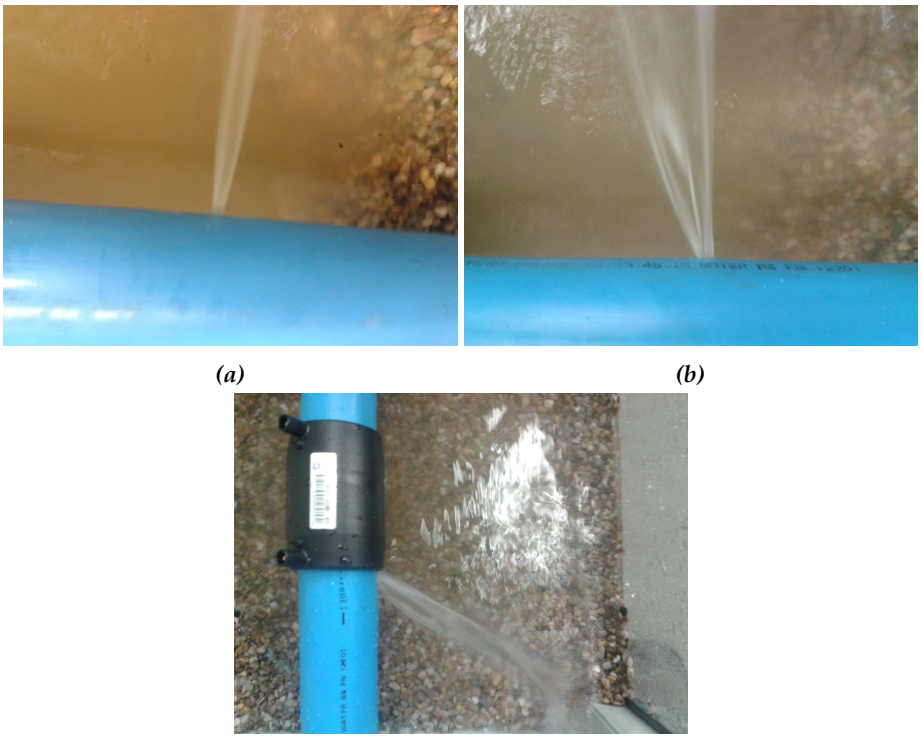


Figure 3.14: *Schematic demonstrating the void created before welding the electrofusion joint.*



(a) *(b)*

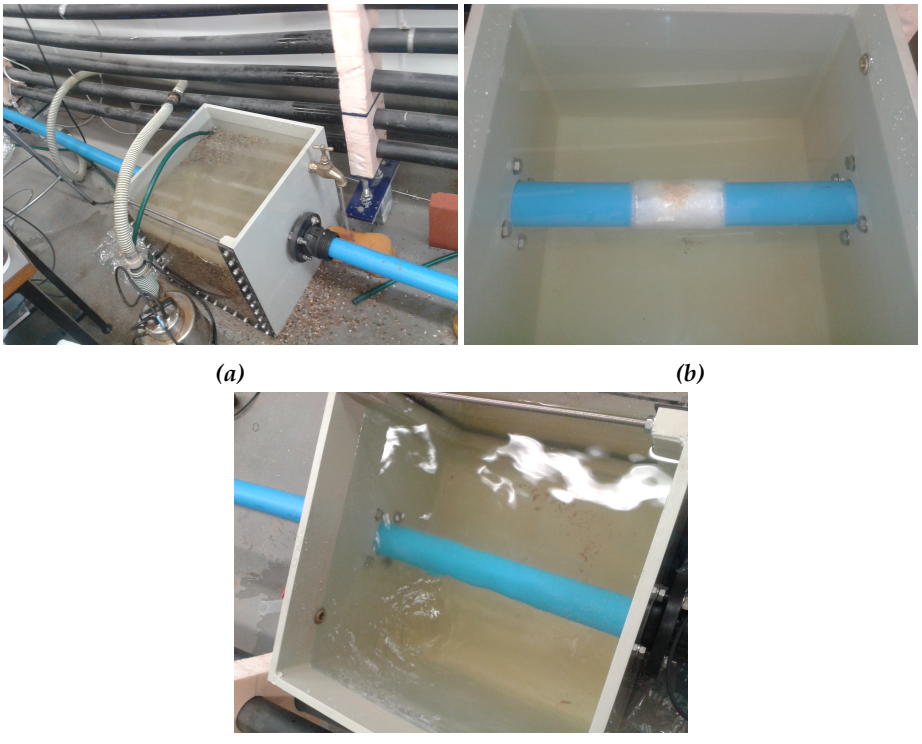
(c)

Figure 3.15: *Photographs of the discharge pattern of leak shapes studied: (a) round hole; (b) longitudinal slit; and (c) leaky electrofusion joint. Note that the gravel has been excavated partly round the leak for the purpose of these photographs.*

3.2.4.3 B: An investigation in to the influence of backfill on leak signals

Changes to the pipe backfill can result in alterations to the leak flow rate (Fox et al., 2014) due to changes in the constraints around the leak. Whilst some studies have attempted to evaluate the effect of backfill (see Muggleton and Brennan (2004) and Pal (2008)), these studies are limited by their experimental design - i.e. they use different pipe rigs for each backfill and therefore the experimental conditions are different for each test. Moreover, if comparing one backfill type to another, there is likely to be a difference in leak flow rate due to some media types being porous and some having a more constrained influence on the leaks discharge (Fox, 2016). Therefore, to fully characterise the effect of backfill the leak flow rate needs to be controlled and kept the same between each backfill type, essentially isolating backfill as the parameter influencing the leak signal.

During experimental test, the test section box was filled with three different types of backfill, beginning with a compact gravel backfill representing a realistic porous media. This was a 5-12 mm diameter pea gravel backfill, in accordance with British Standards for backfill of plastic pipe (BSI, 1973) and therefore represents a standard external media. The box filled with gravel would then be filled with water in order to create a fully saturated gravel backfill. The other two backfill types included a submerged pipe (representing an idealised fluidised backfill) and a geotextile fabric (representing a constrained porous media (Fox, 2016)). Pictures of the backfill types are shown in Figure 3.16. System pressure was adjusted for each backfill type in order to ensure the leak flow rate was the same for each backfill type and therefore the effects of backfill was fully isolated. For each leak shape and leak area studied, the backfill type was changed. Therefore, every leak studied was also studied under gravel, submerged and geotextile fabric backfill types.



(a) *(b)*

(c)

Figure 3.16: *Photographs of leaks in the different backfill types studied: (a) gravel media; (b) geotextile fabric; and (c) submerged.*

3.2.4.4 C: An investigation in to the influence of leak flow rate on leak signals

The influence of leak flow rate was evaluated by adjusting system pressure by turning the downstream gate valve in order to order to increase leak flow rate. This is a common method of leak flow rate adjustment, shown by many authors (Pal et al., 2010; Almeida et al., 2014a). 7-9 different leak flow rates were investigated for each leak area, shape and backfill type given in A and B of this section.

3.2.4.5 D: Summary of data collected

Parts A, B and C above describe unique experiments that, whilst conducted as part of separate experiments for the purpose of isolating individual variables, together the investigations were carried out in order to fully understand the effects of the variables on the leak signal. As a large and complex data set has been collected, the data is summarised in Table 3.3. Each test was repeated 20 times.

Table 3.3: *Summary of the data collected from the pipe rig (No. Q's = number of leak flow rates studied).*

Shape	Width	Length	Diameter	Depth	Area	No.Q's	Repeats
Slit	2mm	5mm			10mm^2	7	20
Slit	2mm	8mm			16mm^2	7	20
Slit	2mm	12mm			24mm^2	8	20
Slit	2mm	16.5mm			32mm^2	7	20
Round hole			3.5mm		9.6mm^2	9	20
Round hole			4.5mm		15.9mm^2	7	20
Round hole			5.5mm		23.8mm^2	8	20
Round hole			6.5mm		33.2mm^2	7	20
Electro Joint	4mm			4mm	16mm^2	7	20
Electro Joint	8mm			4mm	32mm^2	7	20

3.3 Field work experiments investigating the influence of pipe material on the leak signal

A set measurements from real WDS also took place in order to understand the effect of pipe material on the transmission of the leak signal. Details for these tests are shown below.

3.3.1 Experimental design

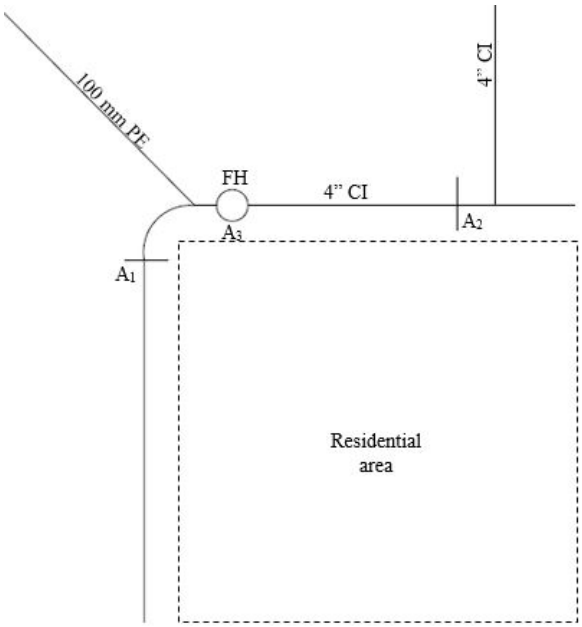
Three different study sites were carefully chosen, all with a different pipe material. A CI, AC, and PE pipe was investigated in order to investigate 3 common pipe materials currently used in WDS (Pal, 2008). To overcome the presence of other physical factors that may influence the leak signal, a novel methodology was designed in order to standardize leak flow rate and leak type, therefore isolating the influence of pipe material on the leak signal. A standpipe was connected to a Fire Hydrant (FH) for all pipe materials, and was used to create an artificial leak. Leak flow rate was controlled by manually adjusting the gate valve at the end of the standpipe and leak flow rate was measured using an electromagnetic flow meter (ABB Aquamaster). Standpipes connecting to FHs are regularly used in order to replicate leaks in WDS (see for example, Almeida et al. (2015a) and Gao et al. (2002)). Three different leak flow rates (0.1 l/s, 1 l/s and 5 l/s) were measured and this was replicated for all pipe materials. A photograph of the standpipe with flow meter and gate valve is shown in Figure 3.17.

Figure 3.17: *Standpipe with flow meter attached to fire hydrant on CI main. The same stand pipe and flow meter was used on all the pipe materials studied.*

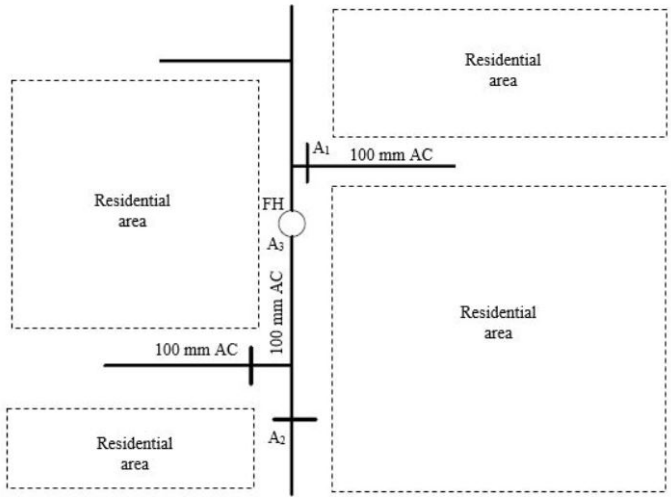
Due to the complexities of real WDS and the required access points, it was not possible to locate study sites which had the same distance between accelerometers and the stand pipe, therefore this varied between study sites. The position of the FH as a leak source relative to the positioning of the accelerometers is given in Table 3.4. System pressure was measured using the closest water company pressure transducers, which reported system pressure of 37.40 m for the CI main, 45.30 m for the AC main and 49.95 m for the PE main. Schematics of the study sites are shown in Figure 3.18.

Table 3.4: *Test site details, where: d distance;* A_1 *Accelerometer 1;* A_2 *Accelerometer 2; and* A_3 *Accelerometer 3.*

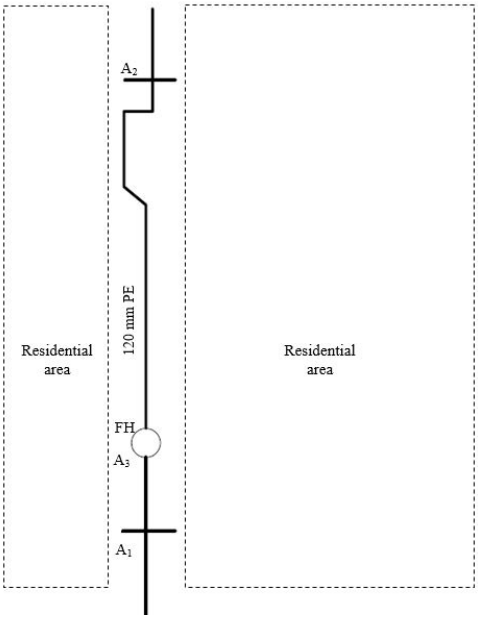
Pipe Material	Diameter (mm)	d leak to A_1 (m)	d leak to A_2 (m)	Total d A_1 to A_2 (m)
Cast Iron	101.6(4")	9.7	35.3	45
AC	100	7.9	17.8	25.7
PE	120	10.1	15.1	25.2



(a)



(b)



(c)

Figure 3.18: *Schematic of the WDS network of each pipe material studied and position of the accelerometers and stand pipe. FH denotes Fire Hydrant, where the artificial leaks were created. (a) CI pipe; (b) AC pipe; and (c) PE pipe. FH Fire Hydrant; A_1 Accel 1; A_2 Accel 2; and A_3 Accel (next to leak). Not to scale.*

3.3.2 Field work experiments: signal and data acquisition

A different accelerometer type was selected for the purposes of the field experiments due to the fact that the accelerometers used in the laboratory study were non-portable, i.e. they lacked sufficient water proofing and also acquired data transfer to a desktop computer, making them impractical for use on real WDS. For the tests on real WDS, three accelerometers were used at a sensitivity 10 V/g to record the leak's VAE signal simultaneously during the field experiments. The accelerometers recorded at a sample of 4864 Hz for 60 seconds. One accelerometer was placed next to the leak, and one accelerometer upstream and another accelerometer downstream. Details of the accelerometer locations and schematics of each pipe networks is provided in Figure 3.18. The accelerometer distances from the artificial FH leak measured using a measuring wheel are provided in Table 3.4. Following the generation of the artificial leak, signals were imported into Matlab for signal processing. A 4^{th} order lowpass Butterworth filter was used to filter signals from all pipe materials. Signals from the CI and AC pipes were filtered <2500 Hz and for the PE main <400 Hz in keeping with common set points in the literature (Pal, 2008). Signals were then segmented into 3 second samples and averaged to remove noise. The soil type was addressed by analysing UK soil types, and found that the soil types for each case was the same, defined as "slowly permeable seasonally wet acidic loamy and clayey soils" (Landis, 2016). The wavespeed of the leak noise was calculated to be 1220 m/s, 1080 m/s and 370 m/s for the CI, AC and PE pipes respectively using the equation given in Almeida et al. (2015a).

3.4 Chapter summary

This chapter introduced the experimental pipe-rig used to study the fundamental factors that influence a leak signal. Moreover, an extensive programme of experiments was carried out in order to commission the pipe rig, which helped to gain an understanding of how the pipe rig operates allowing for an incite in to designing improved experimental methodologies. A unique methodology consisting of several separate testing procedures is proposed which isolates individual variables so that their effects on the leak signal can be fully investigated. The chapter also describes the field work sites and experimental methodology proposed to characterise the effect of pipe material on the transmission of leak signals. The data and understanding resulting from the information provided in this chapter will be used to inform the classification/prediction of leak flow rate, leak area and leak shape presented in later chapters.

Chapter 4

An experimental investigation in to the physical factors that influence leak signals

4.1 Introduction

As discussed in the literature review (Chapter 2), existing investigations in to the factors influencing leak signals have been somewhat limited, due to the inherent variability in signal characteristics caused by lack of control within the experimental methodologies. This is due to the fact that studies do not fully isolate individual variables and therefore their impact on a leak signal cannot be understood. The aim of this chapter is to present some of the results of the experimental investigations detailed in the methodology section (Chapter 4). This section is divided in to two parts, which both utilise unique experimental approaches: (1) laboratory experiments investigating the influence of leak flow rate, leak area, leak shape and backfill type on leak signals; and (2) controlled experiments taking place on real, complex WDS whereby the influence of pipe material (Cast Iron, Asbestos Cement and Plastic) on the leak signal is investigated.

4.2 (a) Laboratory experiments

The simultaneous measurement of leak flow rate, system pressure and leak VAE signal under laboratory conditions was unique to this study. The testing procedure involved the adjustment of system pressure in order to vary the leak flow rate. Moreover, the use of the system 'test section' allowed for changes in leak shape, leak area and backfill type.

4.2.1 Influence of leak flow rate on the leak signal: experimental results

In order to investigate the influence of leak flow rate on leak signals, the system pressure was adjusted in order to increase leak flow rate, beginning with the gate valve fully open and turning 1 whole turn 360^0 clockwise to increase leak flow rate. Between 7-9 different leak flow rates were measured for each sample, that being all leak areas, all leak shapes and all backfill types. Figure 4.1 demonstrate a 4.5 mm round hole leak at 40 l/min in the time domain. The influence of leak flow rate on various hole sizes and shapes was assessed and there was a general increase in leak signal amplitude when leak flow rate increased. However, the demonstration of leak signals in the time domain provides little relevant information about the leak itself.

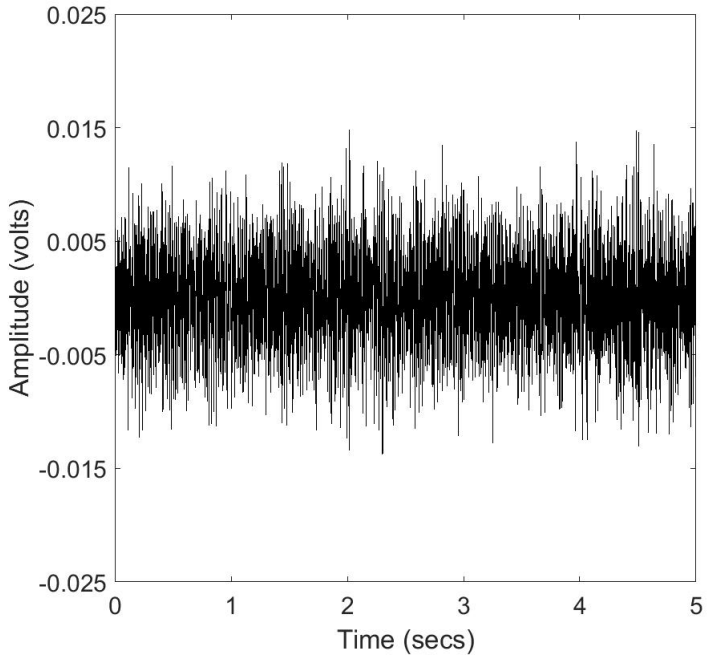


Figure 4.1: *Time domain signal of a 4.5 mm round hole leak at 40 l/min.*

Observations of the time domain spectrum with increasing leak flow rate suggested that the leak signal amplitude increases as leak flow rate increases. However, Figure 4.1 provides little information about the frequency content of the signal, and therefore signals were translated in to the frequency domain. The frequency domain characteristics of round hole leaks of varying diameter (3.5 mm, 4.5 mm, 5.5 mm, and 6.5 mm) at different leak flow rates (varying between 19.03 l/min to 61.77 l/min) are shown in Figure 4.2. Here, the signal is displayed as a function of the ratio of leak:no leak for each hole diameter. 3 leak flow rates are given for each leak size, in which the downstream gate-valve was adjusted in order to increase system pressure and therefore increase the flow rate of the leak.

Plotting the leaks as a function of the ratio of leak:no leak shown in Figure 4.2 allows for the characteristics of the leak signal to be fully established. Background noise was defined as the signal present only in the 'no leak' case and was found (on close examination of the signal) to be at frequencies <28 Hz for all round hole diameters and leak flow rates. The spectral pattern is fairly similar for all leak flow rates, where there tends to be an increase in signal amplitude with increasing frequency up to a point until a plateau is observed. This occurs at all the leak

flow rates analysed. However, the frequency at which maximum amplitude occurs appears to vary, and this seems to be leak flow rate dependent. In the case of the 3.5 mm round hole (Figure 4.2a), the two higher leak flow rates shown (21.83 and 24.51 l/min) appeared to have a very similar spectral pattern, where amplitude increased with frequency to a maximum frequency of 649 Hz and then slowly declined in amplitude until 916 Hz where the signal was mostly compromised of background noise. This suggests that the bandwidth of the leak signal is 28 Hz to 649 Hz in this case. In the case of the 3.5 mm diameter round hole at the lower leak flow rate of 19.03 l/min, interestingly the frequency with the maximum amplitude occurred at 592 Hz. However, there was still a general pattern of increasing amplitude with increasing frequency.

Observations of the other round hole diameters, 4.5 mm, 5.5 mm and 6.5 are displayed in Figure 4.2b, Figure 4.2c, and Figure 4.2d respectively, which demonstrates a similar pattern of increasing amplitude with increasing frequency. In all cases, as leak flow rate increases there is an increase in signal amplitude. The spectral pattern is also of interest within a specific frequency band, between 500 Hz and 650 Hz. Here, there is drop in amplitude and then a rise which appears to occur at approximately the same frequencies for all of the round hole diameters, at all of the leak flow rates shown.

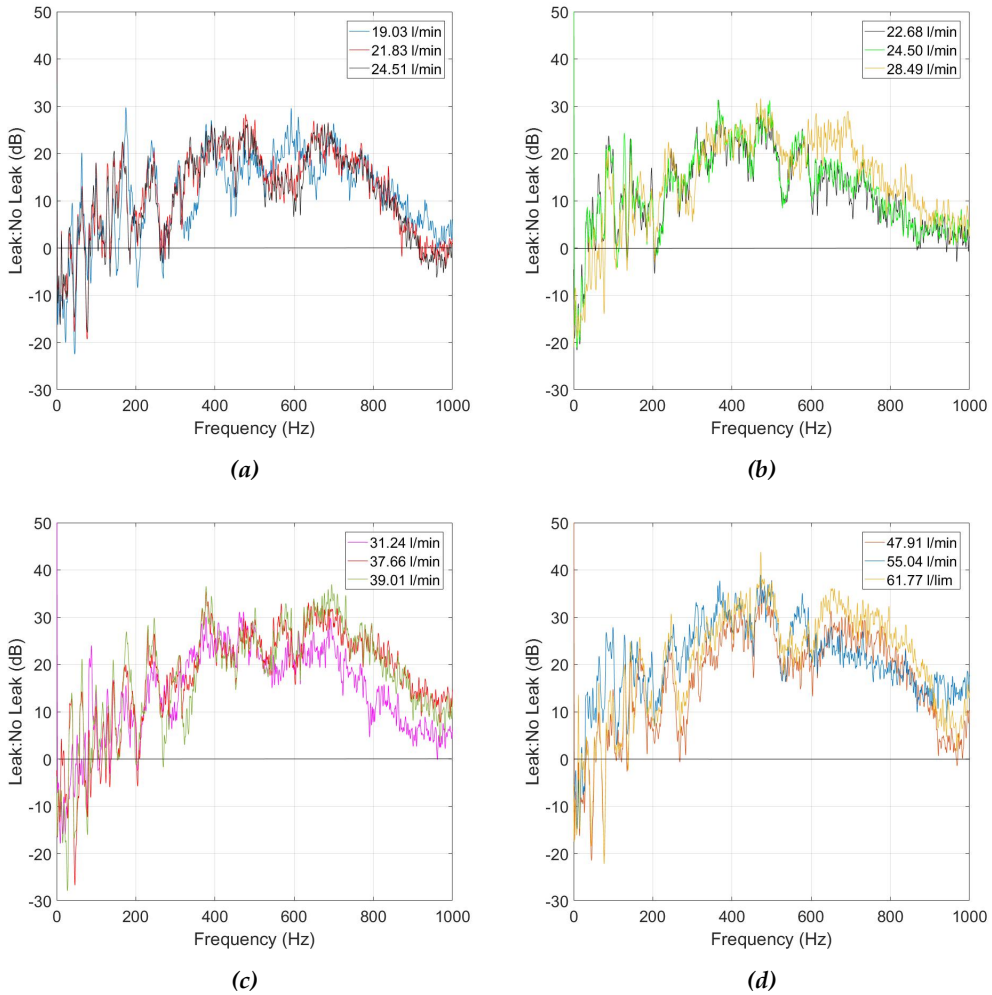


Figure 4.2: *Ratio of leak:no leak for round hole leaks: (a) 3.5 mm; (b) 4.5 mm; (c) 5.5 mm; and (d) 6.5 mm hole diameter.*

The response of the longitudinal slit to increased leak flow rates was also assessed, and this is shown for several sized longitudinal slits in Figure 4.3. The leak area size increases for each element of the Figure, with Figure 4.3a representing an 5 x 2 mm longitudinal slit, Figure 4.3b a 8 x 2 mm longitudinal slit, Figure 4.3c a 12 x 2 mm longitudinal slit and Figure 4.3d representing a 16.5 x 2 mm longitudinal slit. Similarly as the round hole leak shape shown in Figure 4.2, the longitudinal slit appeared to also increase amplitude with leak flow rate as system pressure increased and this was the case for all the different sizes of longitudinal slits assessed. The maximum amplitude from all the longitudinal slits measurements occurs at 390 Hz with an amplitude of 31.2 dB (Figure 4.3d). Similarly to the round hole leaks shown above, the longitudinal slits also observe a dip in amplitude at around 500 Hz which then recovers at approximately 650 Hz, until the decrease with amplitude and frequency continues. However, this does not occur to the same degree at all leak flow rates. In the case of the medium flow rate for the 8 x 2 mm longitudinal slit (Figure 4.3b), the drop in amplitude is of less significance. In the case of the 12 x 2 mm longitudinal slit (Figure 4.3d), despite a difference in leak flow rate, there is minimal difference to the signal spectra, whereas the other longitudinal slit sizes have more obvious divisions between flow rates shown within the frequency domain.

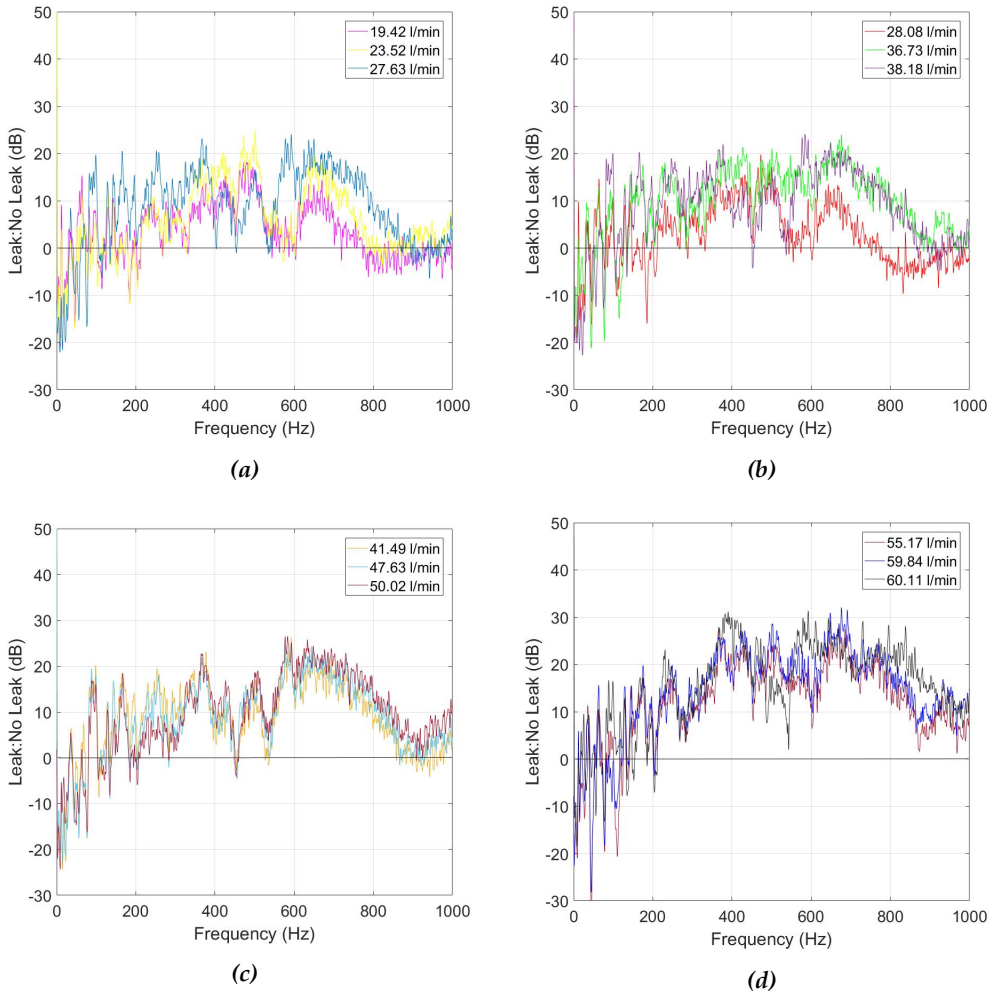


Figure 4.3: *Ratio of leak:no leak for longitudinal slit leaks: (a) 5 x 2 mm; (b) 8 x 2 mm; (c) 12 x 2 mm; and (d) 16.5 x 2 mm.*

Figure 4.4 demonstrates the frequency domain spectra of 2 leaks created by removing a small

void from the pipe material before an electrofusion joint was welded (shown in Figure 3.14). As joint leaks are more common on plastic pipe, this leak has been designed in order to more closely represent 'real' plastic pipe leaks. Figure 4.4a and Figure 4.4b represent the recorded signals of the leaky electrofusion joints of sizes 4 x 4 mm and 8 x 4 mm at different leak flow rates as a ratio of leak:no leak. Coherent with the round hole and longitudinal slit leaks, both of the assessed electrofusion joints also increased signal amplitude as leak flow rate increased. However, the visible separation between leak flow rates was less clear compared to the other two leak shapes, and this was especially the case in the smaller electrofusion joint leak (Figure 4.4a).

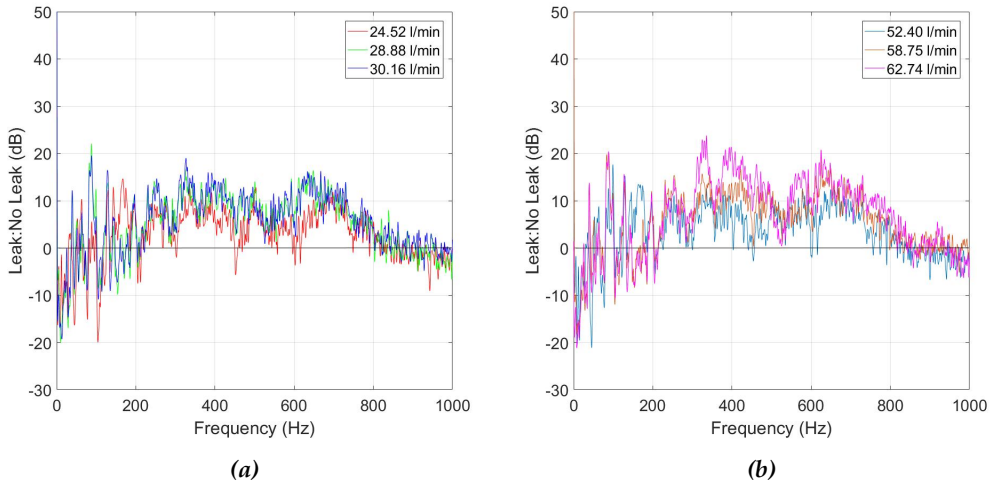


Figure 4.4: *Ratio of leak:no leak for electrofusion joint leaks: (a) 4 x 4 mm; and (b) 8 x 4 mm.*

The figures above have demonstrated that signal amplitude increases with leak flow rate for varying sized round holes (Figure 4.2), longitudinal slits (Figure 4.3) and electrofusion joints (Figure 4.4). However, it is difficult to quantify a whole signal change in amplitude by these frequency domain plots alone. A simple way of assessing the change of amplitude with leak flow rate is by using a single feature which represents signal amplitude, such as the leak signal RMS level. A preliminary investigation into the 3 of the leaks presented in Figure 4.2 at leak flow rates of 37.66 l/min, 47.91 l/min and 55.04 l/min have RMS levels of 2.10 mV, 3.41 mV and 4.6 mV respectively, suggesting that RMS also increases with leak flow rate. Figure 4.5 demonstrates the signal RMS level for all of the leak flow rates studied, including all leak areas, leak shapes and backfill types. The figure clearly demonstrates that signal RMS increases linearly with leak flow rate, suggesting that as leak flow rate increased signal amplitude nearly always seems to increase. This increase in signal amplitude with leak flow rate was further confirmed in the frequency domain plots shown previously in Figures, 4.2, 4.3, and 4.4.

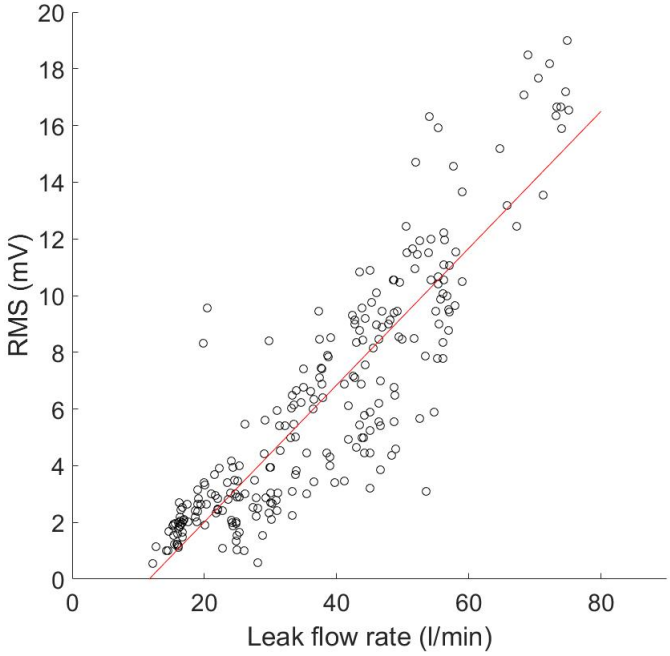


Figure 4.5: *Leak flow rates vs signal RMS, for all leak areas, leak shapes, leak flow rates and backfill types investigated.* R^2*= 0.84.*

4.2.2 Influence of backfill on the leak signal: experimental results

A number of authors have reported that the surrounding backfill can have an influence on the leak signal. However, as the literature review identified, the effect of backfill still remains a major research gap due to the limitations in the existing experimental methodologies presented in the literature. Therefore the true effect of backfill on the leak signal has not yet been quantified. The influence of backfill on the leak VAE signal was experimentally investigated in this study by manually adjusting system pressure and alternating the backfill for each leak shape, leak area and leak flow rate. The unique experimental element of this study was the fact that the system pressure was varied in such a way ensuring that the leak flow rate was kept the same for each backfill type, and therefore the influence of backfill type on the leak signal can be quantified.

The influence of backfill on the leak signal characteristics are shown in the frequency domain in Figure 4.6. Three backfill types are shown (gravel, submerged and geotextile fabric) for a 4.5 mm round hole at approximately 40 l/min. Figure backfill types had highly similar spectrums between the frequency range of 0-590 Hz, whereas the greater differences between backfill types was at frequencies >590 Hz, which showed a reduction in amplitude at these frequencies. As the submerged pipe was a less restrictive material compared to the gravel and geotextile backfill types, the submerged pipe tended to have a slightly higher leak flow rate when the system pressure was consistent across backfill types. Despite this fact, when the gate valve is adjusted to ensure that the leak flow rates are the same for each backfill type, the leak noise is of a lower amplitude in the submerged pipe. For the signals shown in Figure 4.6, the associated RMS values are 4.30 mV, 3.41 mV and 3.99 mV for the gravel, submerged and geotextile fabric backfills respectively. This suggests a slight change in RMS due to backfill type, even though the leak flow rates are standardised. As is the case in the previous examples, the general increase in leak amplitude along with leak flow rate can be shown in terms of changes to RMS level, identified further by labelling the plot by backfill

type. This is shown in Figure 4.7. It is demonstrated that the RMS of leak signals increases despite changes in backfill type, however, in terms of RMS, it actually appears that gravel backfill had a slightly lower RMS compared to the other backfill types, which is not reflected in the frequency domain plots.

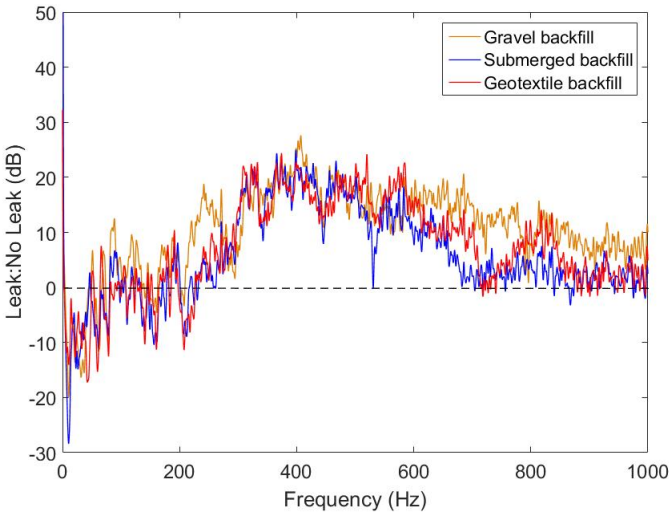


Figure 4.6: *Influence of changing the surrounding backfill on the leak signal.*

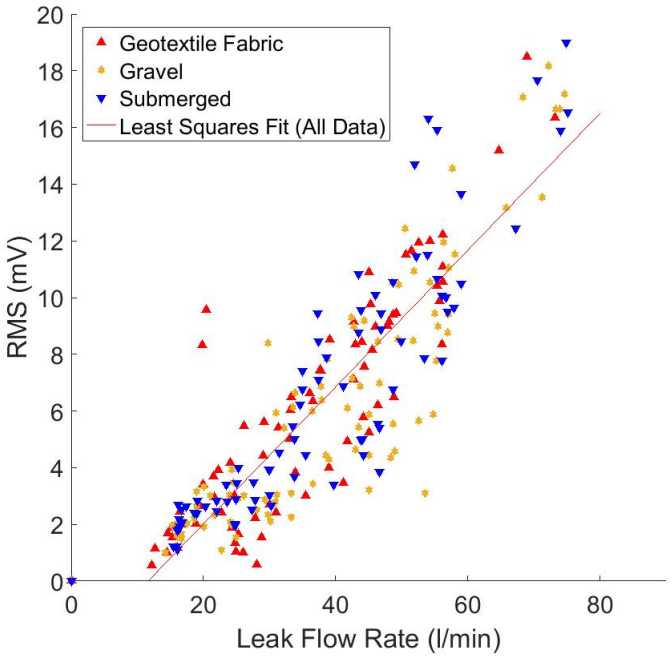


Figure 4.7: *Influence of backfill type on signal RMS at different leak flow rates.* R^2*= 0.84.*

4.2.3 Influence of leak area on the leak signal: experimental results

Intrinsic to this research objective is defining if leak area influences the leak signal. Whilst some studies have investigated leaks of different area by increasing leak area whilst keeping system

pressure consistent, they have not carried out controlled experiments whereby system pressure was varied in order to control leak flow rate. Therefore, the increases in leak area resulted in an increase in leak flow rate and thus the experimental methodologies presented where flawed in this respect as the observed changes in leak signal where likely to be due to the increase in leak flow rate, not leak area. The influence of leak area on the leak signal was investigated in this study by creating leaks of the same shapes but of different leak area (see Table 3.3 which details the size of the leak areas investigated). The downstream gate valve has been adjusted in order to obtain similar leak flow rates, and therefore the influence of leak area can be isolated and the effect of leak area on the leak signal can be investigated.

The resulting frequency spectrum of round hole leaks with the same shape but different leak areas are presented as a ratio of leak:no leak, shown in Figure 4.8. Here, round holes of two different diameters (3.5 mm diameter round hole with area 9.62 mm^2 and a 4.5 mm diameter round hole with area 15.9 mm^2) are presented at two leak flow rates for each leak area (approximately 22 l/min and approximately 24 l/min). As both leak areas have two similar leak flow rates, the effect of leak area is isolated as the effect of leak flow rate is negated. There appears to be a small difference in leak area at the lower frequencies of approximately 121 Hz, although this difference was a slight variation in amplitude and only appeared in the 16 mm^2 area case. Leak area therefore appears to have a negligible effect on the magnitude spectrum of the leak signal, demonstrated by the fact that there is very little difference between signals when the flow rate is kept consistent and compared with different leak areas.

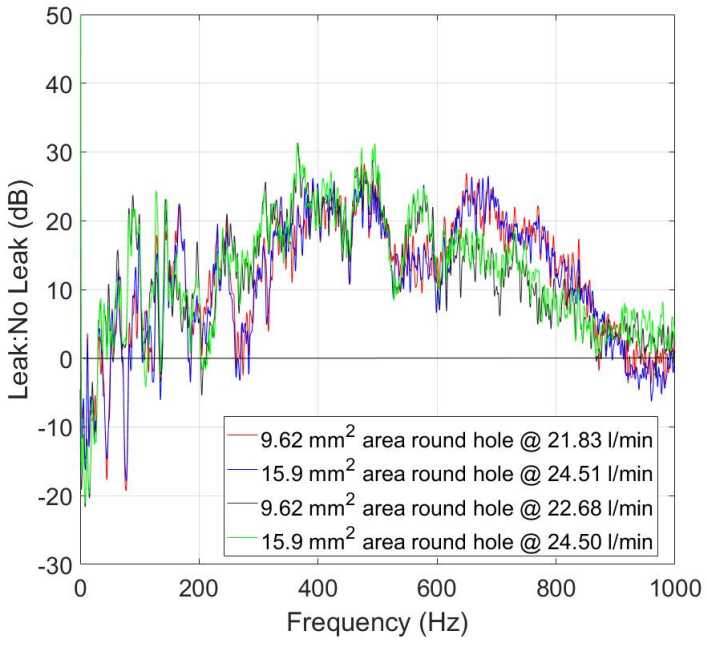


Figure 4.8: *Influence of changing the leak area on the leak signal. Two round hole leaks of area 9.62 mm^2 and 15.9 mm^2 are presented at two different leak flow rates.*

The influence of leak area on signal RMS is displayed in Figure 4.9. The maximum achieved RMS was 19.02 mV, which occurred for the largest leak area (32 mm^2) at the largest leak flow rate. The minimum RMS came from the lowest leak flow rate observed, achieving an RMS of approximately 15.4 mV which was also from the smallest leak area (10 mm^2). The changes to RMS

therefore appear more likely due to leak flow rate rather than leak area, and therefore Figure 4.9 is in support of Figure 4.8 in that the leak area has a negligible effect on leak signals.

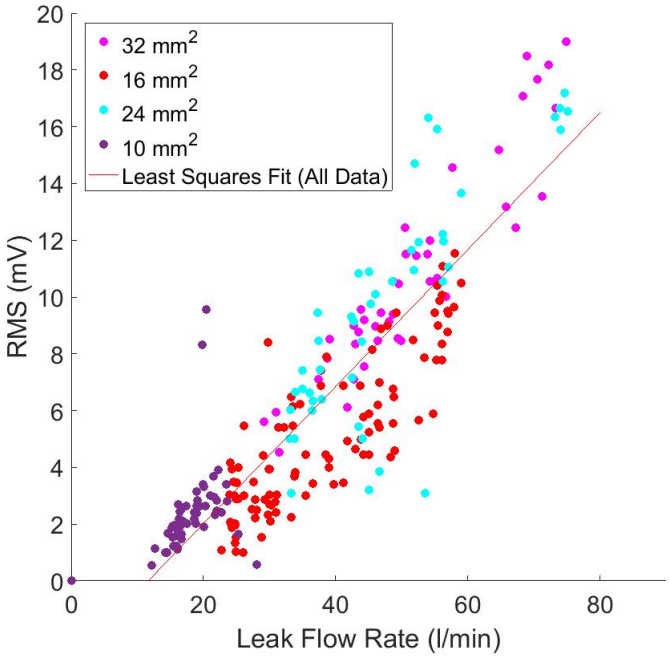


Figure 4.9: *Influence of leak area on signal RMS at different leak flow rates. Colours denote leaks of different area (mm^2). R^2= 0.84.*

4.2.4 Influence of leak shape on the leak signal: experimental results

The literature review also highlighted that leak shape may be a critical factor influencing the leak signal. Once again, studies analysed from the existing literature have not fully isolated the effect of leak shape, and therefore its influence on the leak signal still remains a major research gap. In order to assess the influence of leak shape, leaks of different shape but of equivalent leak area and leak flow rate under the same backfill type were studied. The leak shapes investigated are detailed in Table 3.3, including round holes, longitudinal slits and electrofusion joints. The leak flow rate was again controlled by adjusting system pressure using the downstream pressure using a gate valve, allowing for the isolation of leak shape from other variables.

Figure 4.10a shows a comparison of the leak shapes at a lower leak flow rate of approximately 19 l/min, with both the round hole and electrofusion joint leaks at an equivalent leak area of approximately 10 mm^2. An electrofusion joint was not designed to this leak area size and therefore could not be included in the comparison at this area size. In the case given in Figure 4.10a, both the longitudinal slit and round hole have a similar frequency spectrum at frequencies <145 Hz. At higher frequencies, there is an obvious distinction between the round hole and longitudinal slit obtained through the visual inspection of the signals in the frequency domain. Evidently, the round hole appeared to be higher in amplitude at all frequencies <145 Hz. Moreover, there are distinct differences in signal spectra, where a general reduction in amplitude occurs within the longitudinal slit at frequencies >520 Hz, whereas there is no reduction in amplitude until 592 Hz for the round hole.

In the case of Figure 4.10b, the leak area size has increased and an electrofusion joint measuring 4 x 4 mm is additionally included as well as a round hole (4.5 mm) and longitudinal sit (8 x 2 mm), hence all the leaks shown in this figure have an equivalent leak area of 16 mm^2. Moreover, this figure represents a slightly higher leak flow rate (approximately 28 l/min for all leak shapes) compared to Figure 4.10a. It appears that both the round hole leak and longitudinal slit have a distinct pattern which is similar to the spectrum representing the same shape in Figure 4.10a, indicating that these patterns are representative of these leak shapes. Interestingly, the electrofusion joint had an individual spectrum differing from both the round hole and longitudinal slit, although was similar in spectral pattern at as the round hole and longitudinal slit at frequencies >630 Hz However, the amplitude resembled the longitudinal slit closely at this leak flow rate. The round hole remained as the highest amplitude signal for all the leak shapes in this case.

Figure 4.10c demonstrates a round hole (5.5 mm) and longitudinal slit (12 x 2 mm) at equivalent leak areas of 24 mm^2, at a mid range leak flow rate of approximately 39-41 l/min. Figure 4.10d demonstrates leak area of 32 mm^2, including round hole (6.5 mm), longitudinal slit (16.5 x 2 mm) and electrofusion joint (8 x 4 mm). A higher leak flow rate of approximately 58-59 l/min is also represented. Both of these figures demonstrate that the spectrums shown have a similar spectral pattern at these leak flow rates and area sizes in comparison with the the spectrums shown in Figure 4.10a and Figure 4.10b. However, both figures have slightly higher amplitude, which is most likely caused by the increase in leak flow rate. An interesting comparison lies between the separation between the elctrofusion joint and longitudinal slit in Figure 4.10b and Figure 4.10d, whereby the spectrum appears to be more easily separable between leak shapes at the higher leak flow rate. The figures and description of results have highlighted a strong influence of leak shape on the leak signal. In all cases, round holes tend to have higher amplitude signals. Moreover, the results suggest that each leak shape has an individual frequency spectrum pattern which is recognisable in the frequency domain. Therefore, these results appear to suggest that independent of leak area, backfill and flow rate, leak shape has a significant influence on the leak signal.

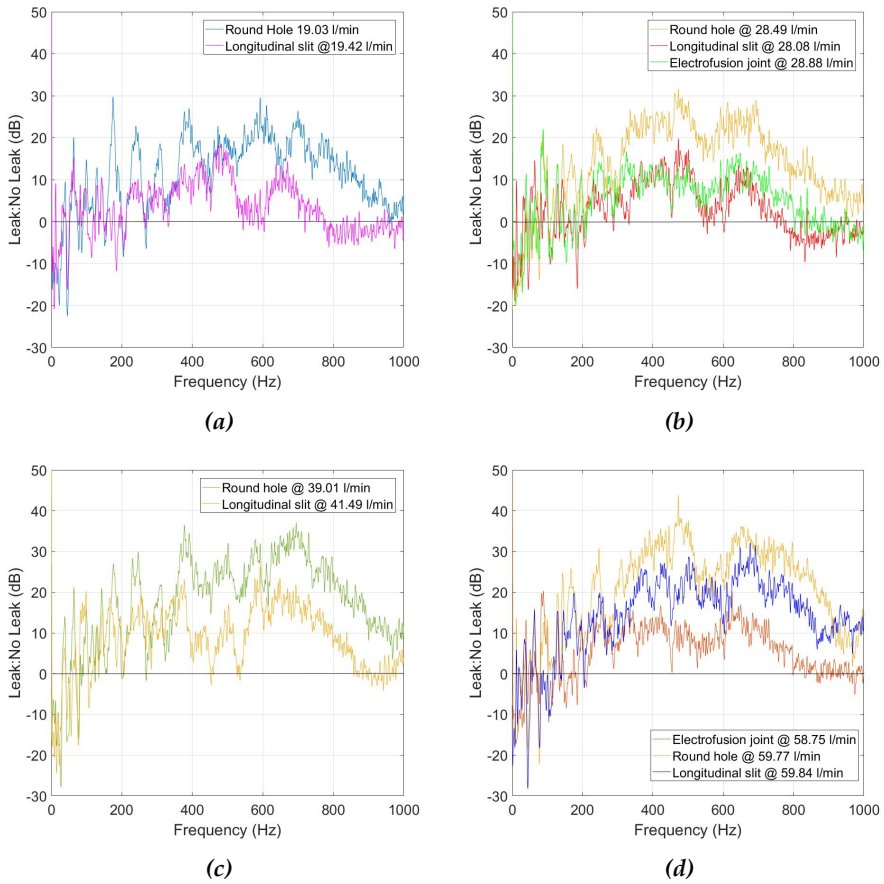


Figure 4.10: *Ratio leak:no leak for different leak shapes at standardised leak flow rates and leak area: (a)Area of 10 mm^2; (b) Area of 16 mm^2; (c) Area of 24 mm^2; and (d) Area of 32 mm^2.*

The influence of leak shape on signal RMS is shown in Figure 4.11. Similarly to the frequency domain plots in shown in Figure 4.10, the round holes appear to have the highest RMS, with the elctrofusion joints having the lowest RMS. However, there appeared to be a smaller difference between signal RMS at the lowest leak flow rates (< 20 l/min). At the medium leak flow rates the differentiation of round holes from longitudinal slits and electrofusion joints is relatively easy using the RMS values. However, RMS alone is not sufficient to distinguish between the longitudinal slits and electrofusion joints at these leak flow rates.

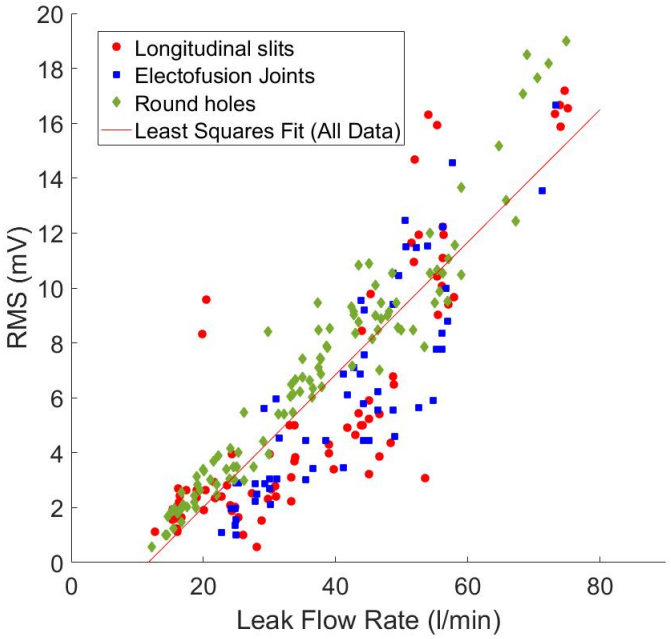


Figure 4.11: *Influence of leak shape on signal RMS at different leak flow rates.* R^2*= 0.84.*

4.3 (b) Influence of pipe material on the transmission of the leak noise: Measurements from real WDS

4.3.1 Introduction

Chapter 2 highlighted that the performance of leak noise correlation has been reported to be highly successful in detecting leaks in metallic pipes, but reports of its efficacy in plastic pipe are varied, highlighting a strong influence of pipe material on the performance of leak noise correlators. However, this has never been experimental quantified in the context of other pipe materials with a good experimental methodology whereby other factors (such as leak flow rate) are standardised. Moreover, the majority of leak detection research using leak noise correlators is conducted on pipe rigs or buried test facilities, which have little resemblance of the complexities of a real WDS. The aim of this section is to study the impact of pipe material on the transmission of the leak signal, investigating different pipe materials in a real, complex WDS.

A rigorous, novel methodology whereby artificial leaks are created using standpipes on different pipe materials and the transmission of the leak noise studied is presented in Chapter 3 (see Section 3.3, "Field work experiments investigating the influence of pipe material on the leak signal"). The results of these tests are demonstrated in the following sections. The effect of pipe material on the transmission of the leak signal was investigated by studying leaks created artificially by opening at standpipe connected to a FH on select sites within real, complex WDS. Three leak flow rates were studied, 0.1 l/s, 1 l.s and 5 l/s on 3 different pipe materials (CI, AC and PE).

4.3.2 Leak spectra measured next to the leak

Initially, leak signal recordings took place by placing the accelerometer next to the leak on the FH spindle. Before any assessment of the leak noise took place, measurements were taken of the background noise to measure the level of ambient noise for each pipe material and this is plotted in Figure 4.12a. These measurements revealed that there were differences in the background noise levels for each pipe, depending on it's location. The differences in background noise may be due to a number of sources as this is based on a real WDS (i.e. background noise could be present due to passing traffic, pipe related noise etc.). However, there were similarities between the background noise at <800 Hz. Unlike in most studies, measurements of the background noise *a priori* allows for a full quantification of the leak signal by taking the ratio of leak:no leak when the leak noise was introduced in Figure 4.12b, Figure 4.12c and Figure 4.12d for the CI, AC and PE pipe materials respectively. Ambient noise appears to dominate at frequencies <10 Hz, where the ratio spectrum is towards 0 dB. It is evident that increasing leak flow rate increased signal amplitude when recorded next to the leak. The pattern was similar in all cases, with an increase in amplitude with frequency until a certain point where this plateaued.

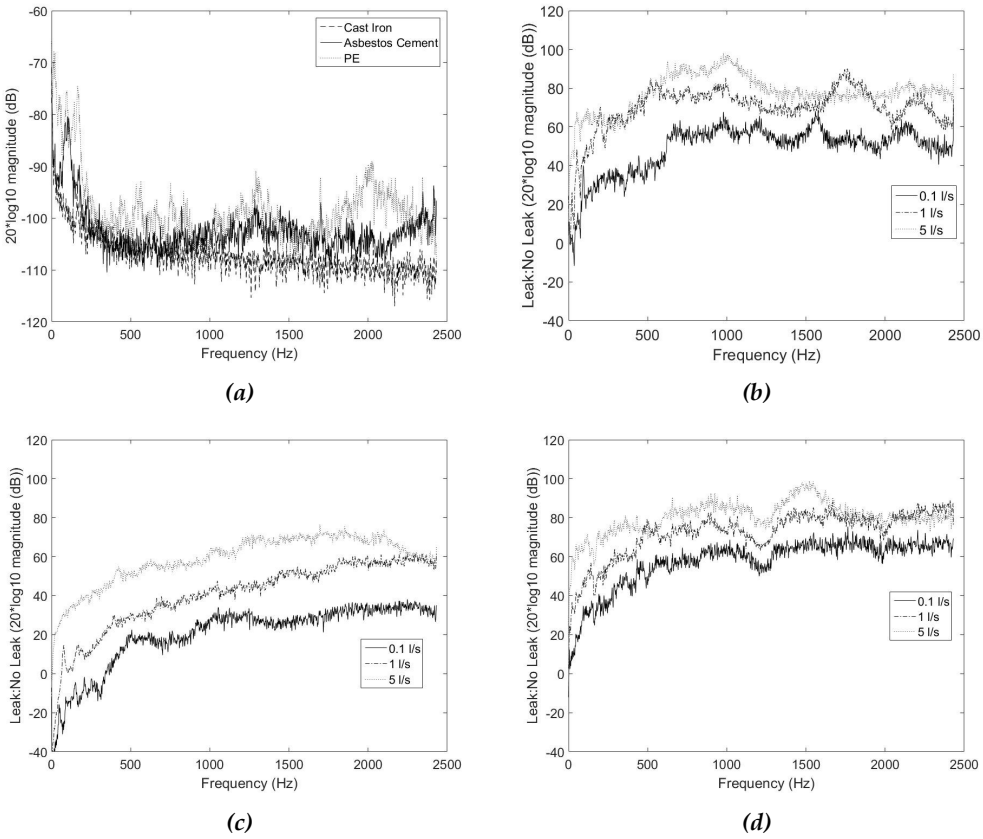


Figure 4.12: *Signal spectra recorded next to the leak for the different pipe materials studied: (a) magnitude of the background noise for all pipe materials; Ratio of leak:no leak for (b) CI; (c) AC; and (d) PE pipe materials. Leak flow rate is standardised to 3 different leak flow rates for all study sites.*

Figure 4.12 clearly demonstrate that the leak signals measured next to the leak are sufficiently similar for all leak flow rates. The measurements taking place in the laboratory shown previously in this chapter have indicated that there is a strong correlation between signal RMS and leak flow rate (Figure 4.5), with measurements recorded next to the leak. This is investigated in the

case of results taken from real WDS in Figure 4.13, whereby the RMS levels at different leak flow rates are compared for the different pipe materials. The figure shows that there is also a strong correlation between signal RMS and leak flow rate. Moreover, similar to the Figure 4.12, there is a strong similarity between signal RMS for each leak flow rate when recorded next to the leak, with negligible differences between pipe materials. Therefore, the effect of pipe material has been isolated and the sole effect of signal transmission can be investigated as signals are sufficiently similar when recorded next to the leak. Any measurements taken further away from this which results in changes to the signal will be due to interactions between the leak signal and some other noise source along the transmission path, and not at the source of the leak.

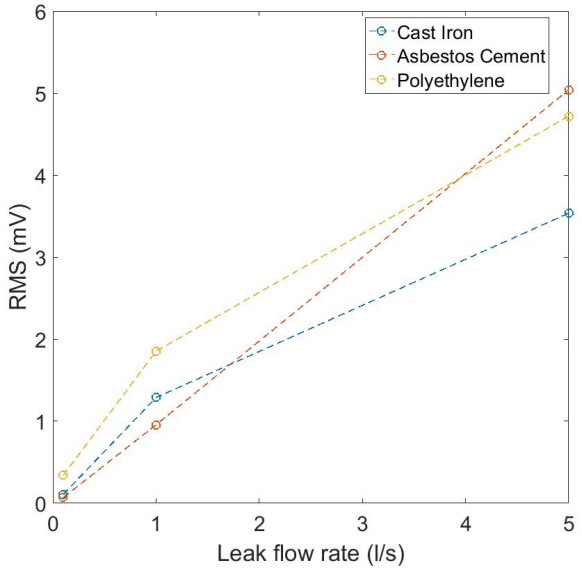


Figure 4.13: *RMS measurements recorded next to the leak at different leak flow rates for the 3 different pipe materials studied.*

4.3.3 Leak spectra measured away from the leak

The influence of pipe material can be assessed by measuring the leak spectra away from the leak, and this is shown in Figure 4.14. The positions of the accelerometer in terms of their distance from the leak is given in Table 3.4 in Chapter 3. As all three accelerometers recorded data simultaneously, the same leak flow rates are recorded. The magnitude spectrum of the background noise (no leak) recorded away from the leak at the downstream accelerometer is given in Figure 4.14a. It was found that there was a slight difference in background noise measuring at a distance away from the leak compared to the background noise levels recorded next to the stand pipe leak (Figure 4.12a). Coherence estimates of the background noise between the two accelerometers positioned either side of the stand pipe leak demonstrate little coherence across all frequency bands (Figure 4.14b, 4.14c, 4.14d). This suggests that the background noise is random and uncorrelated between the sensors.

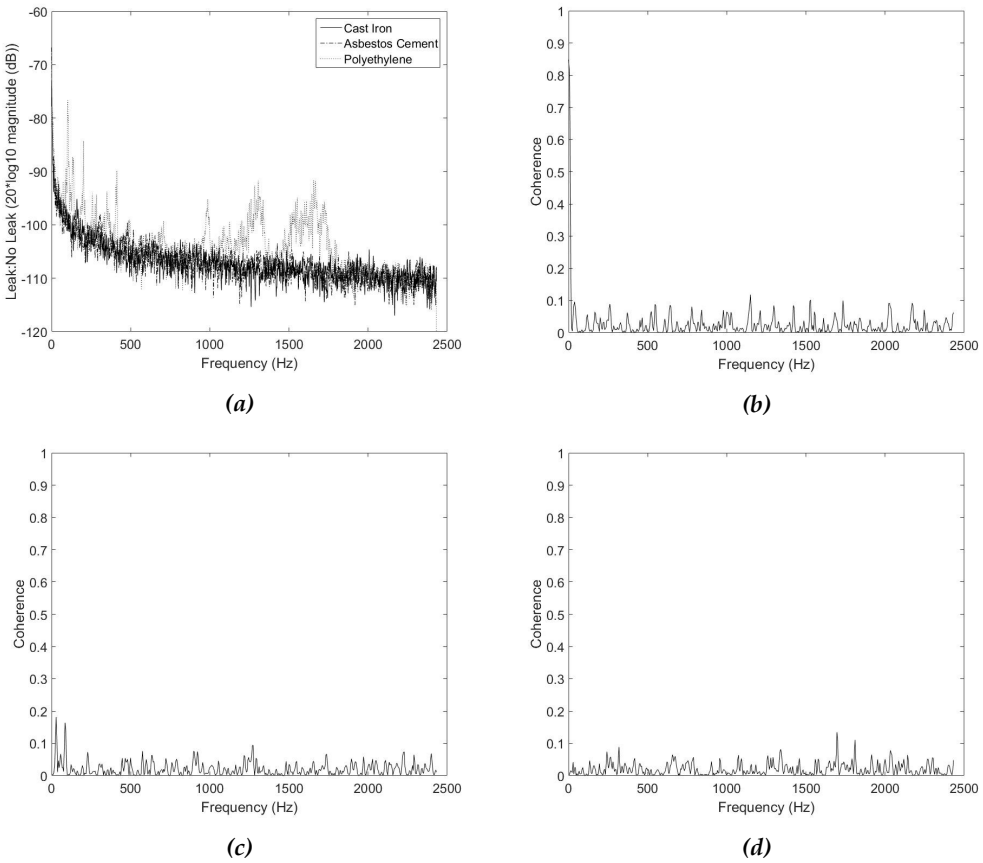


Figure 4.14: *Background noise measurements away from the leak: (a) magnitude spectrum of background noise for all pipe materials; (b), (c) and (d) coherence estimates of background noise for CI, AC and PE pipe materials respectively.*

Turning the stand pipe gate valve to release flow through the stand pipe represented the introduction of an artificial leak. The magnitude spectrum of the downstream sensor and coherence estimates measured at distances away from the leak as a component of leak:no leak are shown in Figure 4.15 for each pipe material. In comparison to the leak spectra measured next to the leak (Figure 4.12), there was a reduction in amplitude at all leak frequencies for all pipe materials. However, this reduction appeared to be less for CI and AC mains compared to the PE pipe and this was the same for all leak flow rates. Both of these pipe materials also demonstrated good coherence within a similar frequency range (approximately 250-2300 Hz). These results indicate that CI and AC pipe materials are effective transmitters of the leak signal. Leak signals recorded for the PE appeared to show little difference between the background noise and the leak signal, suggesting that the PE material is a poor transmitter of the leak signal. However, when the signal from the PE pipe was filtered to <400 Hz, some signal still remained at <256 Hz in the PE pipe at 5 l/s. However, no signal was present at leak flow rates of 0.1 and 1 l/s. It is evident that low pass filtering the signal component for the PE main helped improve the identification of the leak noise, but only at the highest leak flow rate. Little coherence was observed for the PE main, and it was found to have coherence similar to the background noise recordings, demonstrating the attenuative properties of the PE pipe material.

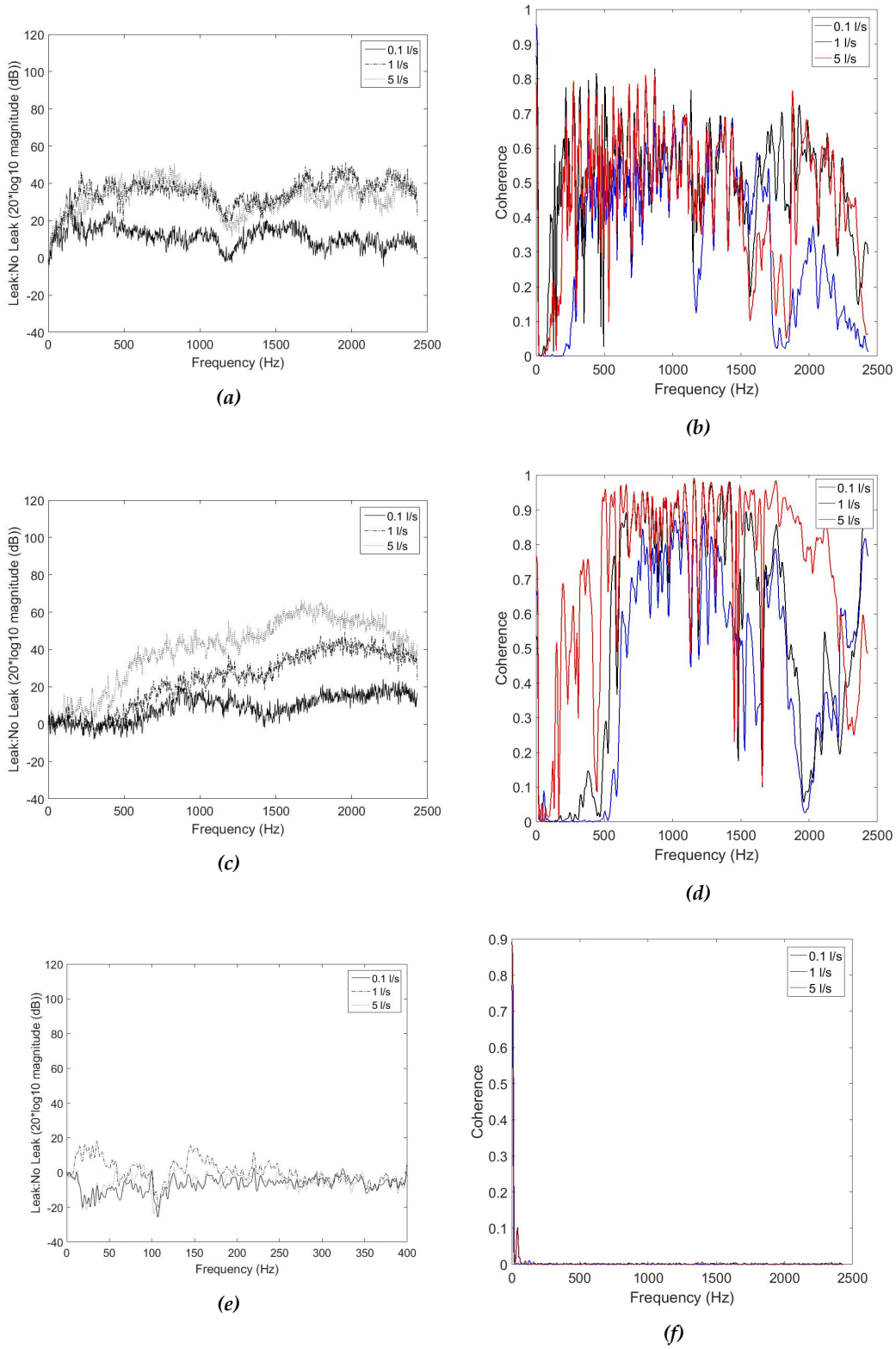


Figure 4.15: *Leak signal measurements recorded away from the leak (downstream accelerometer), magnitude and coherence estimates at 3 different leak flow rates: (a) Magnitude of CI pipe; (b) coherence of CI pipe; (c) magnitude of AC pipe; (d) coherence of AC pipe; (e) magnitude of PE pipe; and (f) coherence of PE pipe.*

The measured vibration signals were also processed for phase information, and the unwrapped phase spectra is shown in Figure 4.16. In the case of the AC and CI mains (Figure 4.16a and Figure 4.16b), there appears to be a largely linear phase-frequency relationship within specific frequency

bands. However, in both of these cases there appears to be numerous distortions in the phase spectra suggesting it is not entirely linear. There also appears to be an effect of leak flow rate on the relationship between phase and frequency. In the case of the CI main, at the medium (1 l/s) and high leak flow rates (5 l/s), the onset of linear phase-frequency is at around the same frequency (approximately 152 Hz for both leak flow rates). In the cases of the lowest leak flow rate (0.1 l/s), linear phase-frequency was not observed until 209 Hz. The AC main also observed the onset of linear phase-frequency at approximately 156 Hz for the leak flow rates of 1 l/s and 5 l/s, whereas for the 0.1 l/s leak flow rate this occurred at 441 Hz, suggesting the leak flow rate has a greater impact on the phase-frequency relationship. In the case of the PE pipe, it is shown that there is no linear phase-frequency relationship (Figure 4.16c).

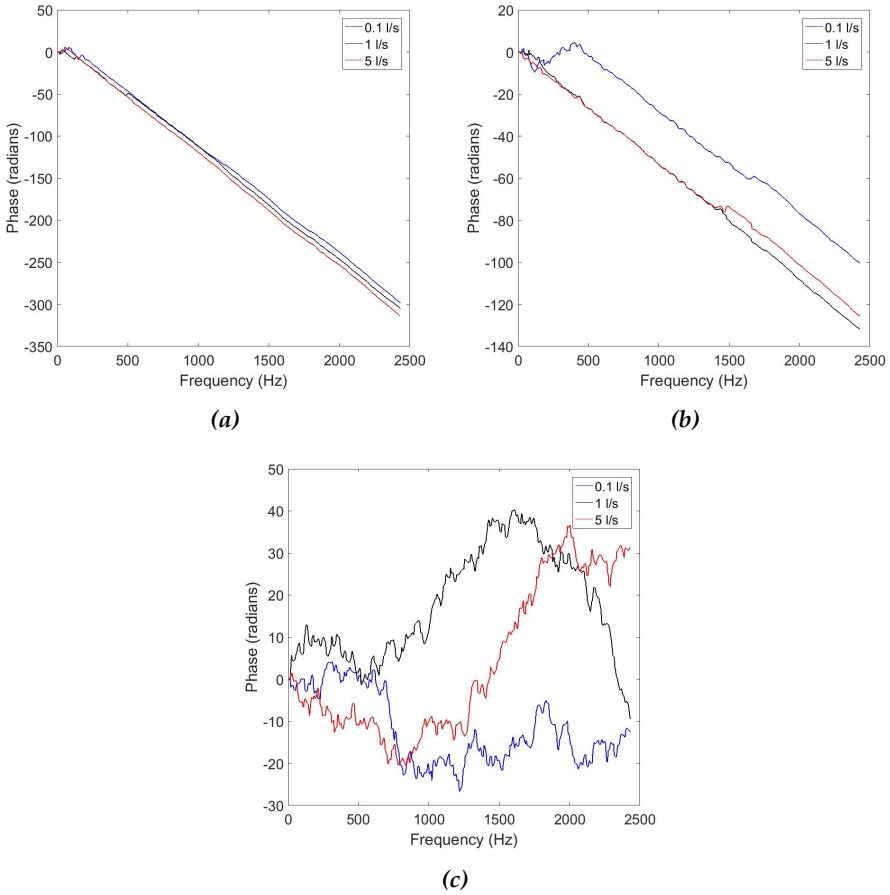


Figure 4.16: *Unwrapped phase spectra at three different leak flow rates for the different pipe materials: (a) CI; (b) AC; and (c) PE pipe materials.*

4.3.4 Influence of transmission distance

The signal attenuation rate for the different pipe materials was calculated from the frequency-amplitude response of the sensor next to the leak and the sensors placed further away from the leak, placed both upstream and downstream. These are plotted in Figure 4.17. The pipe material is shown to influence the rate of attenuation, and this appears to be independent of leak flow rate and frequency for the CI and AC pipe material. Attenuation rates for the PE appear to be more frequency dependent, with the higher frequencies attenuating more quickly compared to the lower frequencies. The numerical attenuation model calculated using Equation 2.8 and Equation 2.9

(described in Chapter 2) showed a fairly good fit at the lowest leak flow rate, but accuracy was reduced at the higher leak flow rates.

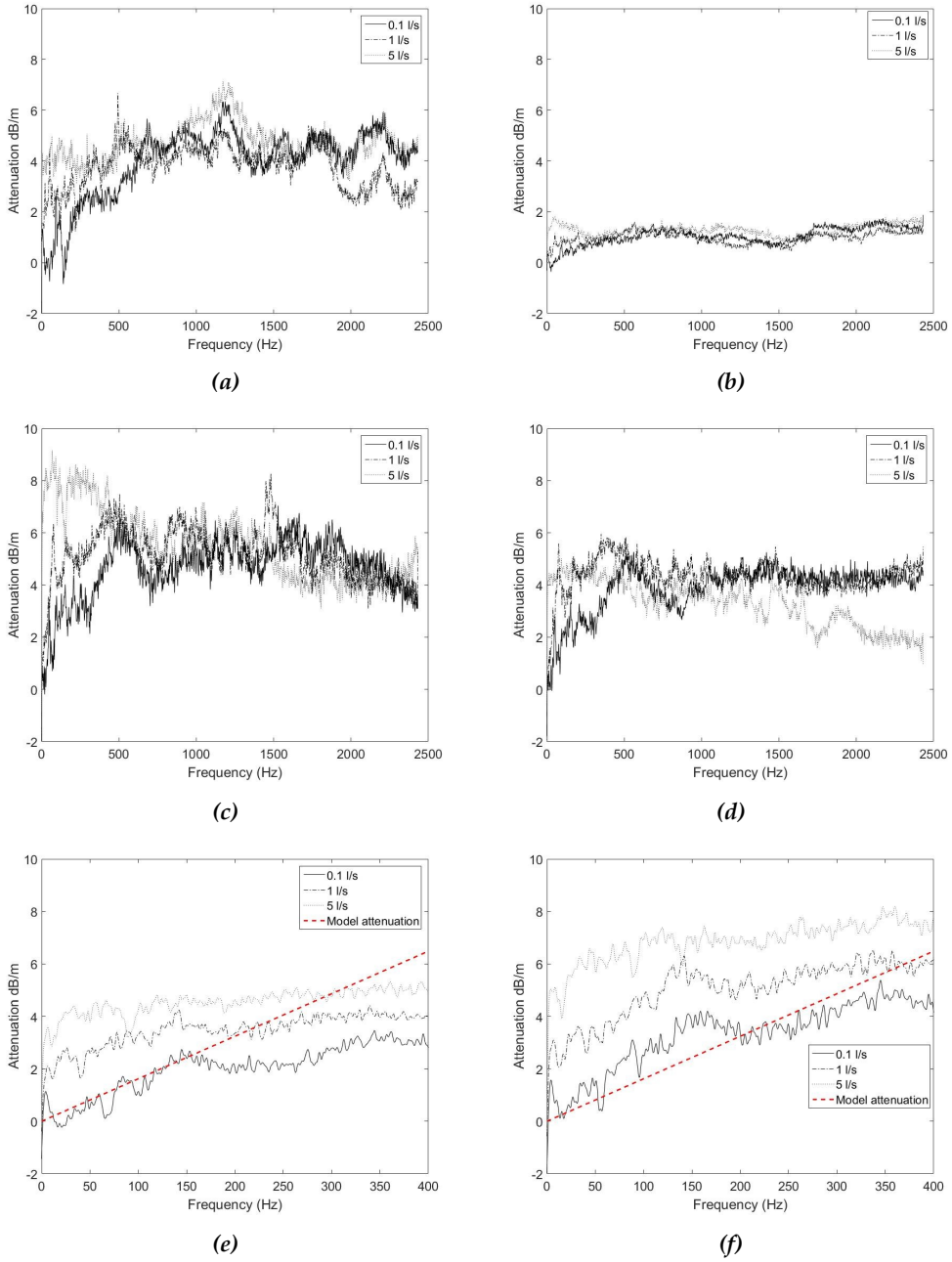


Figure 4.17: *Attenuation plots for the upstream (Accelerometer 2) and downstream (Accelerometer 1) accelerometers. (a) CI (Accelerometer 1 - 9.7 m from leak); (b) CI (Accelerometer 2 - 35.3 m from leak); (c) AC (Accelerometer 1 - 7.9 m from leak); (d) AC (Accelerometer 2 - 17.8 m from leak); (e) PE and (c) PE (Accelerometer 1 - 10.1 m from leak); and (f) PE (Accelerometer 2 - 15.1 m from leak). Note that the plotted signal for the PE is the product of a filter set <400 Hz, whilst AC and CI signals are filtered <2500 Hz.*

Both the upstream (Accelerometer 2) and downstream (Accelerometer 1) accelerometers are shown for all pipe materials, and appear to show a difference between the upstream and downstream attenuation rates. This is especially prevalent for the CI main (Figure 4.17a and Figure 4.17b) which has lower levels of attenuation travelling to accelerometer 2 (upstream) compared to that of accelerometer 1 (downstream), despite the fact that the leak noise travels a shorted distance when travelling to accelerometer 2. This implies that attenuation is not linear with distance, or

the complexity of a real world WDS increases attenuation rates (e.g. from branches, reflections etc.). To understand this further, plots of the leak signal as a function of Accelerometer 1/Accelerometer 2 are shown in Figure 4.18. For the CI pipe (Figure 4.18a), there is little difference between the transmission path of Accelerometer 1 and Accelerometer 2 between 0-1000 Hz and >1700 Hz, suggesting that the increased distance between the leak and accelerometer does not increase attenuation at these frequencies bands. At frequencies 1000 - 1700 Hz the leak signal is affected by the transmission path. A similar pattern was observed for the AC and PE mains (Figure 4.18b and Figure 4.18c respectively) where the change in distance between Accelerometer 1 and Accelerometer 2 did not seem to have a significant influence on the leak signal. This suggests that the main loss in leak signal is primarily between the stand pipe leak and the pipe itself rather than increased distance meaning that stand pipe leaks may not be representative of real leaks.

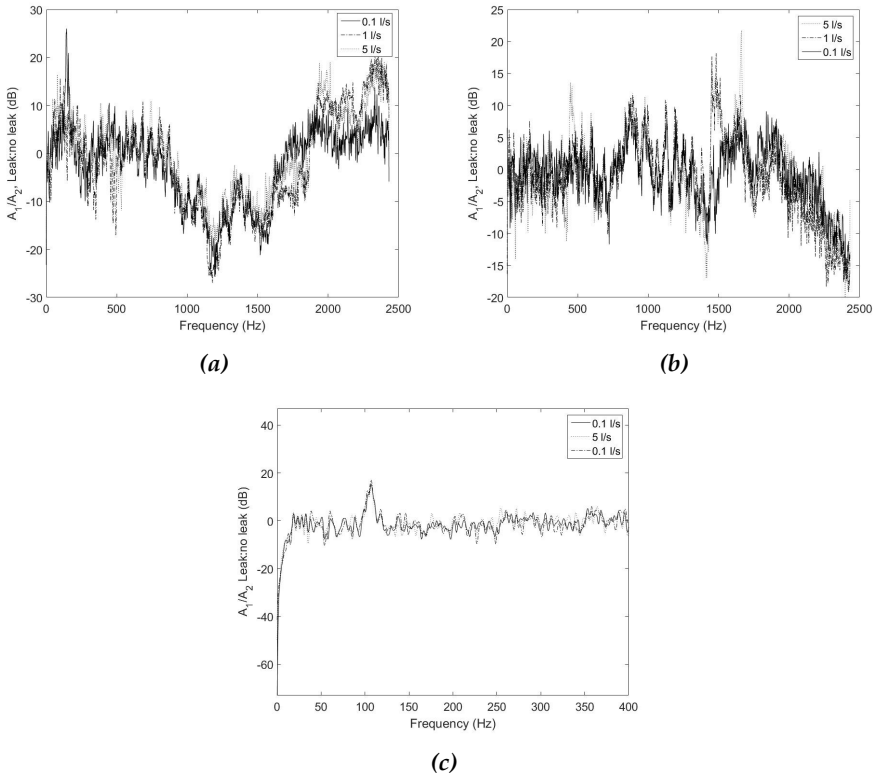


Figure 4.18: *A1/A2 for different leak flow rates in different pipe materials: (a) CI; (b) AC; and (c) PE pipe materials.*

4.3.5 Influence of pipe material on the cross correlation of signals

Leak signals from the three pipe materials have been analysed using the cross correlation method described mathematically in Equations 2.1 to 2.4 in Chapter 2. The literature review (Chapter 2) identified that leak signals can be improved by pre-whitening signals, known as the GCC method. The SCOT estimator and PHAT estimator has been shown to be useful estimators by Gao et al. (Gao et al., 2006). However, of all the GCC filters, the SCOT and PHAT methods appeared to be the most promising when analysing signals recorded on the pipe rig during the commissioning phase (Figure 4.21). The data shown in Figure 4.16 has demonstrated that there is linear phase-frequency relationship at frequencies <110 Hz for the AC, and CI mains, and this appeared to be leak flow rate dependent. In the case of the plastic pipe, there was no linear phase-frequency relationship.

Therefore, the PHAT estimator is unlikely to provide accurate leak location estimates in the case of the PE main and therefore SCOT estimator has been used in order to cross correlate the leak signals for the purpose of this study.

The normalised cross correlation functions using the SCOT estimator (Equation 2.11 and Equation 2.14) are plotted for the different leak flow rates in Figure 4.19, Figure 4.20 and Figure 4.21 for the CI, AC and PE pipe materials respectively. Figure 4.19 suggests that the CI pipe had distinct clear peaks in the cross correlation at all leak flow rates. The τ_{delay} for this leak was estimated to be -0.0094s,-0.0092s and-0.0092s for leak flow rates of 0.1 l/s, 1 l/s and 5 l/s respectively. Leak flow rate therefore appears to have a negligible effect on the time delay estimate for the CI pipe material and leaks in this material can produce strong, distinct, clear peaks in the cross correlation.

The cross correlation of leak signals recorded from the also AC main demonstrated distinct, clear peaks (Figure 4.20). However, in the case of 0.1 l/s, a few peaks are located next to the main peak which could be due to a number of sources, including reflections, discontinuities in the pipe, fittings etc. As the signal becomes attenuated with distance, cross correlation could be an issue in this case as the peak for the leak would be less clear and distinguishable compared to the other the other peaks. Increasing leak flow rate in this case appears to make a more distinct peak. Calculated τ_{delay} for this leak are -0.0170s, -0.0086s, 0.0084s for leak flow rates of 0.1 l/s, 1 l/s and 5 l/s respectively. The difference in τ_{delay} at 0.1 l/s is of significance and provides a different leak location estimate compared to the other two leak flow rates, therefore it is likely the main peak was not the peak due to the leak in this case. In comparison with the other two pipe materials, it was not possible to cross correlate the leak noise for the PE main and no distinct peak was identifiable to the leak was found, despite there being some identifiable leak signal at <256 Hz.

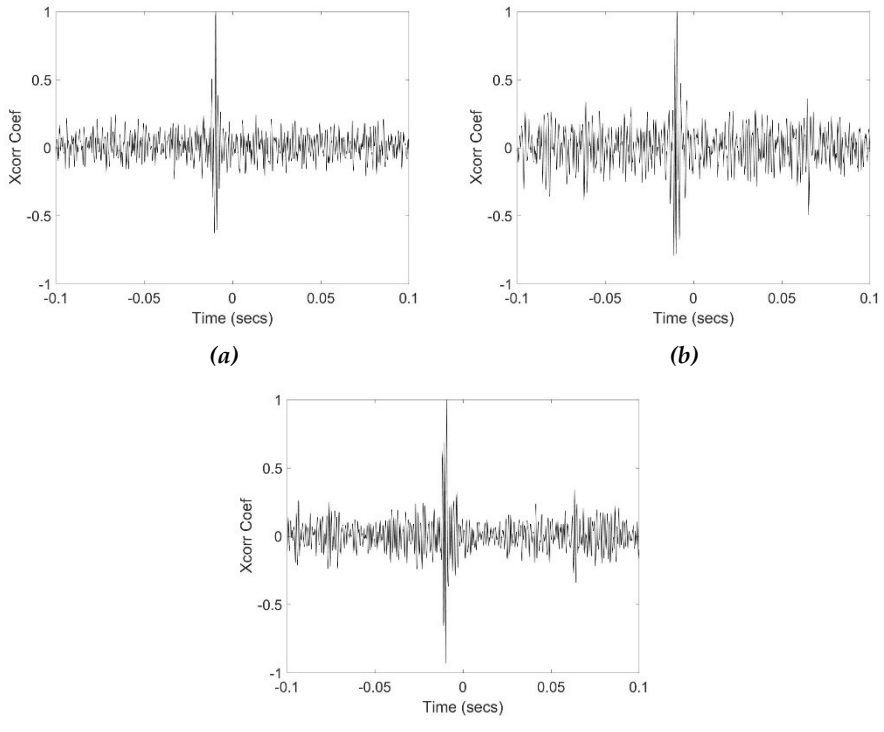


Figure 4.19: *Cross correlations using the SCOT estimator for the CI pipe at the different leak flow rates: (a) 0.1 l/s; (b) 1 l/s; and (c) 5 l/s.*

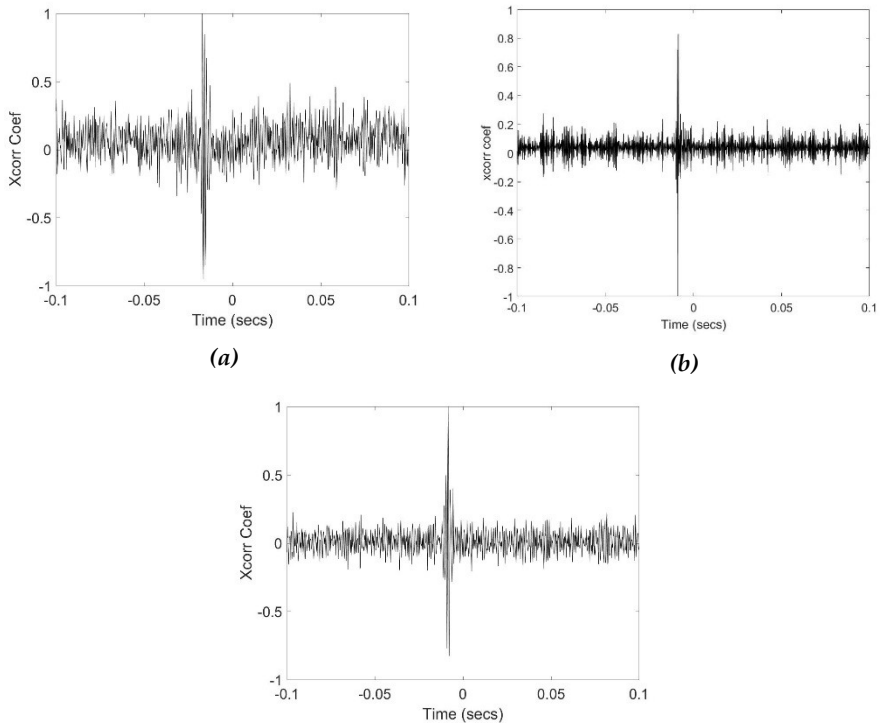


Figure 4.20: *Cross correlations using the SCOT estimator for the AC pipe at the different leak flow rates: (a) 0.1 l/s; (b) 1 l/s; and (c) 5 l/s.*

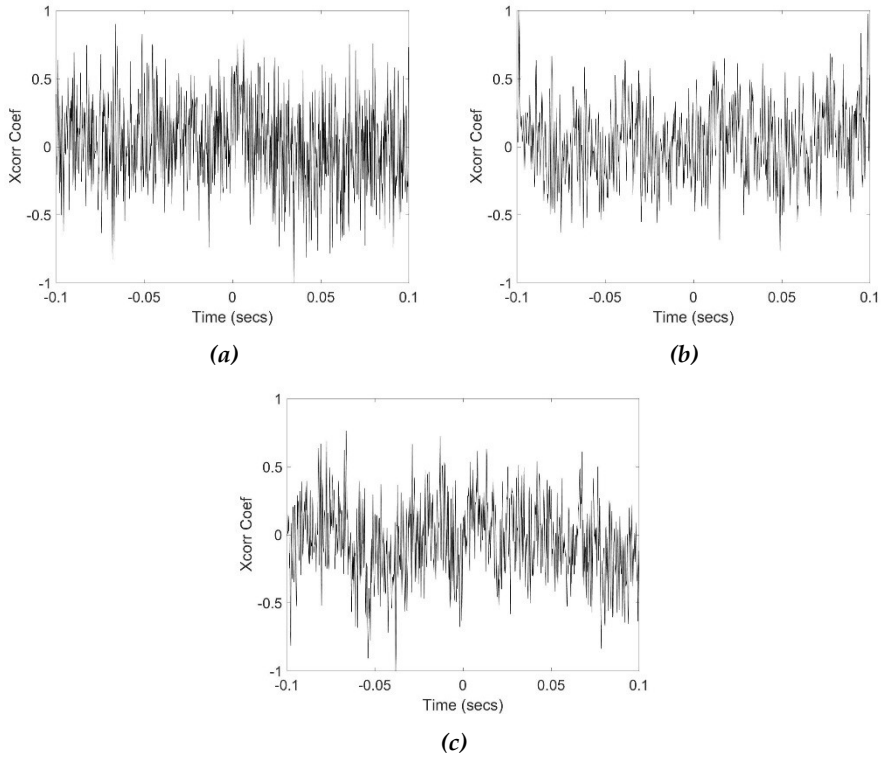


Figure 4.21: *Cross correlations using the SCOT estimator for the PE pipe at the different leak flow rates: (a) 0.1 l/s; (b) 1 l/s; and (c) 5 l/s.*

4.4 Discussion around the factors influencing leak signals

4.4.1 Introduction

The work presented herein investigates some of the fundamental factors that are thought to influence leak signals but still remain not fully understood within the literature, highlighting a major research gap. This took place in the laboratory investigating the influence of leak flow rate, leak area, leak shape and backfill type on leak VAE signals. Additional experiments also took place on real WDS from a UK water company, investigating the influence of pipe material on leak VAE signals.

4.4.2 Discussion of the laboratory investigations

A unique methodology was used whereby the system pressure was adjusted in order to standardise parameters, thereby isolating variables and therefore their effect on the leak signal could be understood. This involved adjusting the downstream gate valve to vary system pressure, creating situations where leak flow rate, leak area, backfill and leak shape was isolated individually.

4.4.2.1 Discussion on the influence of leak flow rate

It can be seen from the results presented that leak flow rate had a strong effect on the leak signal. Increasing leak flow rate led to an increase in amplitude of all frequencies greater than the background noise level (>28 Hz) (Figure 4.2). This was consistent in the cases of all leak shapes (Figure 4.2, Figure 4.3 and Figure 4.4), leak are and backfill types. The observed increase in signal amplitude is coherent with other studies (for example Pal et al. (2010)) and it is also likely that the higher leak flow rate leaks will be more easily identified than those with smaller leak flow rates (Humphrey et al., 2012) due to the increased signal amplitude. The plots demonstrating changes in signal amplitude through RMS values and leak flow rate furthermore reflected increases in signal amplitude with increases in leak flow rate (Figure 4.5). This is coherent with the studies by Chen et al. (2007) and Kaewwaewnoi et al. (2010) who found strong correlations between signal RMS and leak flow rate. Leak flow rate therefore appears to have a strong influence on the leak signal characteristics and it seems that there may be enough information within the leak signal to establish models to predict leak flow rate.

4.4.2.2 Discussion on the influence of leak area and leak shape

Anecdotal evidence suggests that larger area leaks make less noise, despite the fact there is no evidence for this. An investigation in to the effect of leak area on leak signals is shown in Figure 4.8, which suggests that there are small differences in leak area at lower frequencies (Figure 4.8) although this is only slight. These differences are more likely due to experimental features, such as the cutting process and localised material stresses (Fox et al., 2016) caused when drilling the leak holes. It is conceivable that limitations in the experimental design led to changes in turbulence around the leak as the water jet discharges hole (possible due to the presence of swarf during the drilling process). As turbulence around the leak hole is a strong parameter governing the leak signal (Papastefanou, 2011), it is important that any swarf etc. are either removed or standardised

between studies. However, the presence of swarf in the experimental study was limited by repeatedly inserting and removing the drill bit. Cassa and van Zyl (2011) and Ferrante (2012) have shown leak area to be a key variable in defining the leakage behaviour, however this study has shown that there was a negligible effect of leak area on leak VAE signals, especially in comparison to the effect of leak flow rate established previously. Leak area therefore appears to have a negligible effect on leak VAE signals, and leak flow rate is a more important parameter.

Leak shape was found to influence leak signal amplitude and frequency when the leak flow rate was consistent across all leak shapes, and this was found to be true for all the leak areas studied (Figure 4.10). Although there has been limited investigation in the literature, this study is coherent with some studies in that the leak shape influences the leak signal (for example Pal (2008)). Across the whole signal spectrum, each leak shape has a distinct pattern recognisably associated with it's individual leak shape. However, when a number of leak flow rates are included together, it becomes difficult to differentiate between leak shapes. Differing signal spectra due to changes in leak shape may be due to varying jet angle (Ferrante, 2012) as the leak discharges the hole. In turn, this will likely create varying turbulence regimes around the leak hole specific for that leak shape, where the signal is created (Papastefanou, 2011).

This studied attempted to create leaks that would be representative of real leaks, however, in real world WDS, longitudinal slits are not formed artificially by removing material from the pipe, they are formed through chemical action degrading material (Duvall and Edwards, 2011), fatigue (Nishimura et al., 1993) and slow crack growth (Brown, 2007). The artificial removal of pipe material to create the different leak shapes could have created swarf which may have varied between the test cases, therefore this would create variations in the leak signal. Again, the presence of swarf or any extra material made during the cutting process, as is the case for all tests, may have an effect on the turbulence regime and therefore the leak noise. This study has experimentally determined that leak shape is another key variable in determining leakage behaviour.

4.4.2.3 Discussion on the influence of backfill

The influence of backfill on the leak signal was investigated in Figure 4.6 and Figure 4.7. Evidently, the backfill type has a strong influence on the leak signal and this has also been shown by a number of other authors (see for example Pal (2008), Muggleton et al. (2002) and Muggleton and Brennan (2004)). However, all previous studies are limited in their experimental design as they compare situations where more than one variable is changing (e.g. leak flow rate differs between backfill types), and therefore it is not possible to quantify whether the changing signal is due to a change in backfill or another variable. A good example of this is through the results shown in Pal (2008), who found the frequency range of leak signals in buried MDPE pipe to be in a different frequency range from those in unburied pipe. However, as Pal (2008) notes these variations could be due to the fact that the experimental rig changes from a buried pipe-rig to an unburied pipe-rig (i.e. a different rig is used in both cases), and therefore the pipe, leak and experimental set-up are completely different. There is also a likely alteration with leak flow rate when the pipe is laid under backfill due to media restrictions around the leak. A similar scenario is shown in Muggleton and Brennan (2004) who compared alternate backfills and the influence on the leak signal. However, they used two different pipe rigs buried in two different backfills in order to assess this. Therefore, the effect of backfill has never been fully quantified. By keeping leak flow rates consistent for different backfill types, the

effect of backfill was isolated in this study and this study is the fist of its kind to demonstrate the influence of backfill fully, in a case whereby the backfill has been fully isolated.

This study found that the backfill type had an influence on the spectrum of leak signals. This was shown through differences between backfill type in the frequency domain and through fluctuating RMS values when the backfill type is changed. However, the degree to which backfill influenced leak signals depended on the backfill type and the signal frequency. The submerged pipe tended to have the lowest amplitude signals compared to that of the geotextile and gravel media. This result differs to that of Muggleton and Brennan (2004) who found higher amplitude signals with flow discharging from a leak under a submerged pipe compared to a pipe buried in sandy soil. According to their study, leak energy radiated into soil more so than a submerged pipe, which was not the case in this study. However, the experiments conducted by Muggleton and Brennan (2004) were conducted under two separate pipe rigs, with different leak flows and leak geometry and therefore the test results are not directly comparable. This shows the importance of only changing one parameter when doing this type of experimentation. Moreover, it was found that the backfill had little influence on the signal at frequencies <590 Hz, in agreement with Brennan et al. (2006a) who found that backfill had little influence on leak signal frequencies <500 Hz. The variation in signal RMS observed for different backfill types is possibly due to the fact that some of the leak energy is absorbed by different media particles (Pal, 2008), reducing the amplitude of some frequencies for different backfill types.

4.4.3 Discussion on the field work experiments investigating the influence of pipe material

Knowledge of the effect of pipe material on the transmission of the leak noise on real WDS is currently lacking study, despite received wisdom that the pipe material is a major factor influencing the leak signal. Although there have been some studies investigating the influence of pipe material on a leak signal (Pal, 2008), each study is individual and therefore it is difficult to draw comparisons between them. Moreover, the majority of previous research has been carried out on simple test rigs and does not compare to the signals obtained from real WDS. This study has presented a novel methodology whereby important factors contributing to variations in leak signal, such as leak flow rate, are kept consistent between each case study and therefore the material effects of the pipeline can be isolated and investigated. This study found that when creating artificial leaks using standpipes, a flow meter can be used to ensure leak flow rate is the consistent between studies. Signals measured at the source of the leak were sufficiently similar to determine that this study has isolated the impact of the pipe material on the leak signal. This research therefore has presented a method to experimentally quantify the effect of pipe material on the transmission of the leak signal, measuring leaks on CI, AC and PE pipe.

Background noise was shown to vary depending on the study site. It is unlikely this is related to the pipe material itself, and more likely due to other possible causes of noise (e.g. passing traffic, customers, turbulence, pipe fittings etc.). Any pipe system contains noised which are intrinsic to that pipeline and this varies between pipeline systems (Khulief et al., 2012), and therefore any plotted leak signal will be a component of the background noise and leak. However, changes in background noise and any intrinsic pipe noises have been overcome in this study by plotting the

ratio of leak:no leak.

It was found that the pipe material has a strong influence on the leak signal. The CI and AC mains appear ideal materials for cross correlation due to their high transmissive properties, where signals attenuated less and coherence was good between sensors either side of the leak. When signals on PE pipe were measured 10 m (accelerometer 1) and 15.1 m (accelerometer 2) from the leak, little leak signal remained. However, the use of a lowpass filter (cut off frequency <400 Hz) led to the identification of a leak at 5 l/s at <256 Hz, but not at the lower leak flow rates (0.1 l/s and 1 l/s), highlighting the importance of the correct choice of bandpass filters. This indicates that increasing leak flow rate is more likely to enhance leak identification on PE pipes, and close attenuation should be paid to the correct use of filter cut off frequency. Despite the visibility of the signal in both upstream and downstream sensors when using a low pass filter for the PE pipe, it was not possible to achieve a reliable cross correlation using the SCOT estimator at relatively short correlation distances. The SCOT estimator was previously reported by Gao et al. (2006) to increase the sharpness of correlation peaks, but this method did not help to identify the leak on the PE pipe in this study. These results differ from a number of studies, such as Almeida et al. (2014a) who managed to correlate a leak and identify its location accurately over a total correlation distance of 50 m on plastic pipe. It is likely that the poor correlations on the PE is due to the high level of signal attenuation through damping in the pipe wall and signal radiation to the surrounding media (Almeida et al., 2014a).

It is has been reported by previous authors that signal attenuation is frequency dependent for buried plastic pipes (Muggleton et al., 2004) and buried metallic pipes (Long, 2001). However, this study highlights little frequency dependent attenuation for the CI and AC pipe materials, but the PE pipe material was found to have a frequency dependent attenuation rate, especially at higher frequencies. There appeared to be differences between the upstream and downstream attenuation rates, suggesting that the location of the sensor can have an impact on the measured signal. Plots of A1/A2 in Figure 4.18 suggest that the propagation distance does not influence the signal for all materials at specific frequencies, and this is also uninfluenced by leak flow rate. Therefore the majority of attenuation is most likely occurring when between the connection of the FH and the pipe. Further investigation of this may lead to a better accelerometer placement regime.

The study presented here which compares different pipe materials found that linear phase with frequency occurred within frequency bands for the CI and AC main, and there was no linear phase-frequency relationship observed within the PE pipe material at any frequencies. This study finds differing results compared to that of Pal (2008), who found a linear phase-frequency relationship within the frequency range 20Hz to 250Hz on plastic pipe. However, similarly to Pal (2008) this study found numerous 'ripples' in the phase behaviour which may be related to pipe resonances. Almeida et al. (2014b) also found that pipe line resonances can cause ripples and phase shifts, which will affect the time delay.

The minor differences between leak flow rates and the onset of linear phase revealed in Figure 4.16 suggested that leak flow rate is a factor in determining the onset of the linear phase-frequency relationship, with the onset of linear phase with frequency occurring at slightly lower frequencies when the leak flow rate is higher (1 l/s and 5 l/s), compared to the lower leak flow rate (0.1 l/s). These results are of significance as phase is often used to calculate pipe wavespeed (Pal, 2008) (Gao et al., 2002), and these results show that for CI and AC pipe materials, the leak flow rate is

of importance to the calculation (as linear phase-frequency is required in order to calculate the wavespeed). Moreover, in the cases of using correlation 'pre-whitening' methods such as the PHAT estimator (see Chapter 2), only information relating to the signal phase is used to calculate the time delay (Almeida et al., 2014b) and therefore at the lower frequencies the PHAT estimator will become a less effective method. In terms of the calculation of t_{delay}, the bandwidth containing linear phase-frequency relationship holds the information regarding signal arrival times (Almeida et al., 2014b) and therefore these results are of high importance. This suggests that there will be a slightly higher bandwidth to correlate signals with at higher leak flow rates. Band pass filters can be applied within the same bandwidths of linear phase-frequency in order to improve the accuracy of the t_{delay} estimate, withholding the frequencies of interest (Gao et al., 2004). However, in the case of the plastic pipe, bandpass filtering is unlikely to improve the location estimate.

The strong differences between this case study on real WDS and literature background most likely lie with the fact that this study takes place on real WDS. The majority of previous research takes place on test rigs or isolated buried pipelines, and therefore lack the complexity of a real world WDS. There is a significant lack of literature which studies the performance of leak noise correlators on real world WDS. A number of factors in a real WDS could have contributed to the poor correlations in this study, including previously unknown discontinuities in the pipeline, reflections, changes to surrounding media etc. It may be the case that in order to identify the leak on the PE main, the sensors need to be situated closer together, and it is possible that this leak would go undetected by acoustic methods in plastic pipes (especially at the lower leak flow rate). A report by Humphrey et al. (2012) reported detectability limits of accelerometers, finding that leak flow rates of 3.6 m^3/hr (1 l/s) should be detectable at a distance of 130 m from the leak source on plastic pipe. However, in this study, at a total correlation distance of 25.2 m, it was not possible to correlate the leak despite the application of more advanced signal processing techniques. Analysis of the network schematic for the PE main suggests that there is minimal possible causes of attenuation (i.e. there are no valves or fittings along the pipeline). However, there are some household customers, but these are also present in the other study sites.

There are other possible factors present which may have influenced the cross correlation, such as transducer sensitivity, which varies depending on manufacturer (Humphrey et al., 2012). Humphrey et al. (2012) also reports that for the same leak flow rate, the use of hydrophones significantly increases the chances of detection and therefore the use of other methods and sensors may have improved the detection of the leak in this study. Similar results were found by Gao et al. (2005), who found that, although accelerometers generally have better coherence over a wider bandwidth, hydrophones may have proven to be more effective in this study as the leak noise is more easily detected over long distances using hydrophones compared to accelerometers. However, hydrophones are invasive and are therefore not preferred by Leakage Technicians. Other methods such as GPR and free swimming methods may be less sensitive to changes in material, and therefore these may be useful on PE mains, or maybe a combination of several techniques.

4.5 Chapter summary

This chapter has presented findings whereby some of the fundamental factors that influence leak signals are studied. A novel experimental methodology is presented whereby variables can be

isolated in order to investigate how they effect leak signals. In addition, the transmission of the leak noise for different pipe materials was assessed at different leak flow rates using data collected from a real, complex WDS. The following conclusions can be made from the results presented in this chapter:

- Laboratory investigations have found that leak flow rate, leak shape and backfill type all have an effect on the leak signal frequency and amplitude. Leak area has a negligible effect on leak signals.
- Signal RMS appears to be a good feature describing some of the variable changes.
- Investigations taking place on real, complex WDS showed that the isolation of pipe material from other factors (importantly leak flow rate and leak shape) allows for the study of leak noise transmission on different pipe materials.
- The attenuation of the leak noise on the PE pipe was so rapid on real WDS that much of the leak noise could not be identified. However, a low pass filter allowed for the identification of the leak noise, but it was still not possible to correlate this leak to identify the location of the leak between the two sensors.
- The rates of attenuation recorded from experimental data are much higher than previous laboratory of rig based studies and higher than the theoretical predictions. However, the field studies showed that the theoretical attenuation rates are suitable at lower frequencies for simple, linear pipe networks, although this was not the case at higher frequencies.
- The cross correlations performed demonstrated sharp peaks and good correlations for the CI and AC main, whilst it was not possible to correlate the PE main.

The implications of these results help to better understand leak noise correlation and their performance in real WDS. Leak noise correlation was found to be highly successful on the CI and AC main, but could not identify a leak on the PE main at a relatively small distance for all leak flow rates. Therefore methods of leak location using VAE based methods are relatively ineffective and therefore research into the development of other methods which show promise (such as GPR) may be effective. However, it is possible that with more sophisticated signal analysis, higher sensitivity accelerometers, alternative vibro-acoustic sensors or an enhancement of the leak noise the leak may still be identified via leak noise correlation.

Chapter 5

Leak flow rate prediction

5.1 Introduction

The observations made in Chapter 5 showed that leak signal characteristics vary depending on leak flow rate, leak area, leak shape and backfill type. Of particular importance was the fact that increasing leak flow rate resulted in an increase in signal amplitude and caused small variations in frequency. However, this became more complex when assessing a combination of changing variables, including leak area, leak shape and backfill type. A key objective of this research is to develop a method in order to estimate the leak flow rate from the VAE signals. The aim of this chapter therefore is to develop a model that can predict leak flow rate based on the leak's VAE signal.

Leak flow rate prediction may be possible through the development of models. The models can be directly comparative whereby new data is compared to existing data and similarities found, or predictive whereby a model makes predictions based on patterns recognised from the past data. Of greatest interest to this research is the development of predictive models, whereby predictions of leak flow rate can be made using the existing and new datasets. Additionally, a great advantage of predictive models is that predictions can be made outside of the range of the training dataset, which is especially useful when recording signals in a pipe-rig where only a limited range of leak flow rates can be created. Essential to any of these models is the derivation of a number of parameters that describe the leak signal and enable leak flow rate prediction. These parameters that describe leak flow rate can be termed signal 'features', which are single values describing a specific signal or element of the signal (Boostan and Moradi, 2003). The signal features investigated can be divided in to three sub-bands:

- Time domain features which have simple interpretation and are easy to extract;
- Frequency domain features, obtained from the conversion of the time domain signal in to the frequency domain via the Fourier Transform; and
- Time-frequency based features, obtained using alternative signal processing algorithms. These algorithms are helpful in describing signals that vary in both the time and frequency domains.

A leak flow rate prediction model can be constructed to utilise these features. Two types of model are investigated in this chapter, including: (1) models using a single signal feature; and (2)

multiple feature models which combine a number of features together. Multiple feature models have more in common with machine learning practices. The models used to develop a leak flow rate prediction model in this study are based on information presented in the Literature Review (Chapter 2), and include the following models:

- Single feature linear and non-linear regression.
- Multiple feature models consisting of (1) Multiple Linear Regression (MLR); (2) Multiple feature non-linear regression (MNLR); (3) Partial Least Squares Regression (PLSR); Kernel Partial Least Squares Regression (KPLSR); and Least Squares Support Vector Machines (LS-SVM).

Furthermore, an overview of the model types used and processing steps undertaken is given in Figure 5.1. This chapter initially investigates single feature models and then moves towards an investigation of multi-feature models.

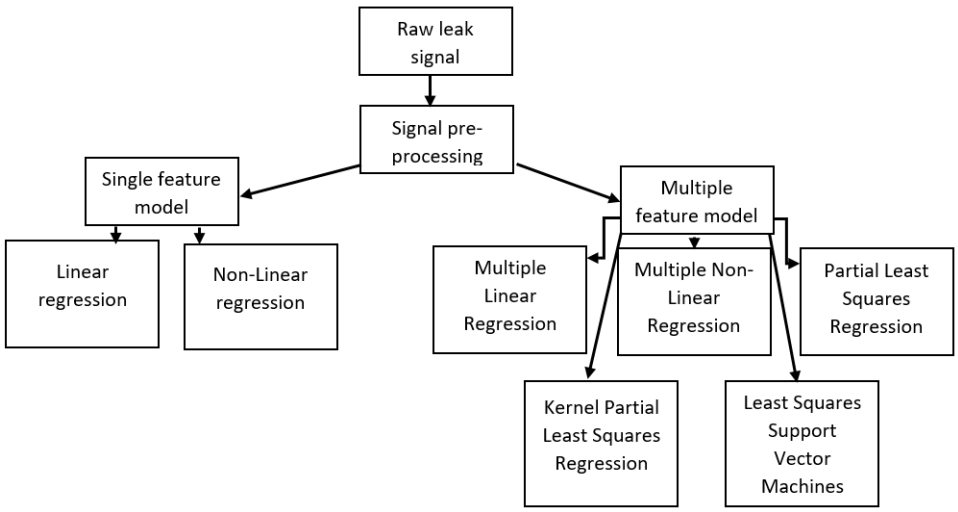


Figure 5.1: *Overview of the model types investigated in order to predict leak flow rate.*

5.2 Data inputs

Following the literature review, a number of signal 'features' were extracted from the leak signals, and these were based on the assessed features outlined in the literature review (Chapter 2). These signal features have previously used by other authors to describe leak signals or to achieve a similar aim. The only inputs in to the models relied on either one or all of the features for the leak flow rates studied, i.e. the models were not told anything else about the leak (such as it's flow rate, shape, backfill or area). For simplicity, the features used are labelled in Table 5.1 and a number is allocated to each feature. The feature numbers are used throughout the rest of the chapter. In the case of single feature models, only one feature is used, but in the case of the multiple feature models all 22 features are used to develop the model utilised.

5.3 Single feature models

As a number of features have been calculated, the most valuable features for use in a single feature model will be the optimum feature describing the variation in leak flow rate. An assessment of

Table 5.1: *Features used as inputs in to the models.*

Features No.	Feature name
1	RMS of whole signal
2-7	RMS IMF1-IMF6
8-13	Spectral [Shannon] Entropy of IMFs 1-6
14	Spectral [Shannon] Entropy of whole signal
15	Mean dB of PSD
16	Maximum dB of PSD
17	Minimum dB of PSD
18	Fundamental frequency
19	Spectral flux
20	Kurtosis
21	Skewness
22	Crest factor

the value of the features shown in Table 5.1 features in describing leak flow rate was made by calculating the Pearson's Correlation Coefficient of the 22 features with leak flow rate. Whilst the Pearson's Correlation Coefficient assumes linearity, it is assumed that this method will still establish whether or not the feature will be useful in predicting leak flow rate. The same approach was also adopted by Chen et al. (2007) in selecting the most valuable features for predicting leak flow rate from gas pipes using acoustic measurements.

The ability of each feature to predict leak flow rate individually can be demonstrated by assessing the relationship between individual features and leak flow rate, and this is given for the features computed in Table 5.2. Here, the Pearson's Correlation Coefficient (Equation 2.30) between leak flow rate and the individual features are shown in the form of matrix. Those features achieving a Pearson's Correlation Coefficient with leak flow rate > 0.75 suggest a strong positive correlation between the signal feature and leak flow rate. This helps to establish any meaningful relationship between features and leak flow rate. It can be seen from Table 5.2 that the most highly correlated variables with leak flow rate (and therefore those that may provide the most valuable features) are the signal RMS and the RMS of IMF1 and IMF2. The other features have a lower Pearson's Correlation Coefficient with leak flow rate. The fact that amplitude dependent features such as signal RMS correlate well with leak flow rate comes is unsurprising as observations made in Chapter 5 demonstrated a number of signals at different leak flow rates which increased in signal amplitude as leak flow rate increased. Further investigation in this section will evaluate the ability of the features highlighted in red within Table 5.2, all other features are ignored for the purpose of single feature modelling. The following sections therefore investigate the use of single feature models as a predictive tool for leak flow rate estimation from VAE signals. The logical first step in developing a leak flow rate prediction model is to therefore utilise features which are shown to have a high Pearson's Correlation Coefficient with leak flow rate (shown by Table 5.2). The strongly correlated features reflected changes in signal amplitude in most cases.

5.3.1 RMS as a single feature model

Signal RMS was found to correlate highly with leak flow rate, achieving a Pearson's Correlation Coefficient of 0.89. The signal RMS is a simple parameter that describes the energy content of a signal and therefore directly reflects changes in signal amplitude (Kaewwaewnoi et al., 2010). RMS

Table 5.2: *Correlation between signal feature and leak flow rate determined using the Pearson's Correlation Coefficient.*

Features No.	Feature name	Pearson's Correlation Coefficient
1	RMS of whole signal	0.89
2	RMS of IMF1	0.81
3	RMS of IMF2	0.79
4	RMS of IMF3	0.51
5	RMS of IMF4	0.35
6	RMS of IMF5	0.19
7	RMS of IMF6	0.13
8	Spectral [Shannon] Entropy of IMF 1	0.13
9	Spectral [Shannon] Entropy of IMF 2	0.48
10	Spectral [Shannon] Entropy of IMF 3	0.28
11	Spectral [Shannon] Entropy of IMF 4	0.03
12	Spectral [Shannon] Entropy of IMF 5	0.04
13	Spectral [Shannon] Entropy of IMF 6	-0.13
14	Spectral [Shannon] Entropy of whole signal	-0.13
15	Mean dB of PSD	0.66
16	Maximum dB of PSD	0.52
17	Minimum dB of PSD	0.11
18	Fundamental frequency	0.03
19	Spectral flux	0.31
20	Kurtosis	-0.03
21	Skewness	-0.00
22	Crest factor	0.41

was calculated for signals at each leak flow rate using Equation 2.28. The maximum RMS value found was that of a 6.5 mm round hole in submerged media, with an RMS value of 18.99 mV at one of the highest leak flow rates of 75 l/min. The minimum RMS value was found for a 8x2 mm longitudinal slit with an RMS of 0.56 mV at one of the lowest leak flow rates of 12.19 l/min. The standard deviation across the whole range of data was 4.3 mV.

A variety of single feature regression model types exist which can be used for leak flow rate prediction, and these can be easily divided in to linear and non-linear models. A comparison was made between linear and non-linear feature modelling techniques and these are given in Table 5.3 where the RMS values for all leak area, shape, backfill and leak flow rates are used as inputs in to the model. The resultant comparison demonstrates that all models provide strong predictive performance with all achieving an R^2 of >0.69 (exponential model). The worst performing model still managed to achieve an R^2 of 0.69 which still implies a positive correlation between leak flow rate and signal RMS using that model type. However, the optimum fit in terms of R^2 comes from using the 4^{th} order Polynomial, achieving an R^2 of 0.80. In terms of model RMSE, the best performance came from using the 2^{nd} order Polynomial with a training and test RMSE of 5.92 l/min and 7.34 l/min respectively, although the R^2 ($R^2 = 0.71$) was slightly less than the 4^{th} order Polynomial ($R^2 = 0.80$) model. A plot of the 2^{nd} order Polynomial fit is provided in Figure 5.2 and the output of the model which assess the models predicted leak flow rate against actual leak flow rate is provided in Figure 5.3. The result shown is the average of the 20 samples taken at each leak flow rate, for all leak area, leak shapes and backfill types. Whilst these results may suggest that the data is more suitable to non-linear models, the difference in terms of RMSE and R^2 is negligible across all models, and as a linear model has the added advantage of simplicity and reduced computational cost, it is deemed that a linear model should be used for the case of predicting leak flow rate with

RMS values (i.e. there is negligible advantage of using a no-linear model in this case).

Table 5.3: *Assessment of the single feature RMS based model to predict leak flow rate, comparisons of linear and non-linear kernels.*

Model	Training RMSE	Test RMSE	R^2	Equation (Q =)
Linear	6.29	7.57	0.78	$3.23mV_{RMS} + 17.26$
Poly 2^{nd} order	5.92	7.34	0.79	$0.08mV_{RMS}^2 + 4.58mV_{RMS} + 13.54$
Poly 3^{rd} order	6.03	7.40	0.80	$0.02mV_{RMS}^3 - 0.49mV_{RMS}^2 + 7.54mV_{RMS} + 8.55$
Poly 4^{th}order	5.92	7.36	0.80	$-0.00mV_{RMS}^4 + 0.06mV_{RMS}^3 - 0.99mV_{RMS}^2 + 9.67mV_{RMS} + 5.98$
Logarithmic	6.49	8.09	0.76	$17.99ln(mV_{RMS}) + 9.07$
Exponential	7.79	8.10	0.69	$19.63e^{0.09mV_{RMS}}$
Power	5.99	7.44	0.77	$14.87mV_{RMS}^{0.52}$

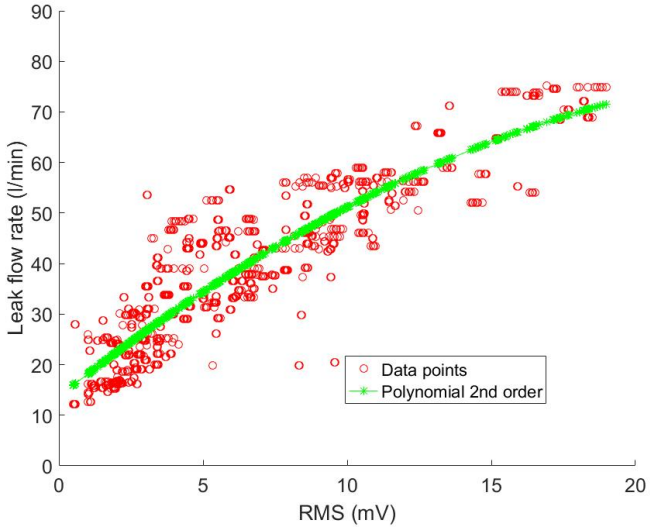


Figure 5.2: *Leak flow rate vs RMS for all leak area, shapes and backfill types (average RMS value of the 20 samples per leak flow rate studied). The data is fitted with a 2^{nd} order polynomial curve. This model achieved an RMSE of 5.92 l/min and an R^2 of 0.79.*

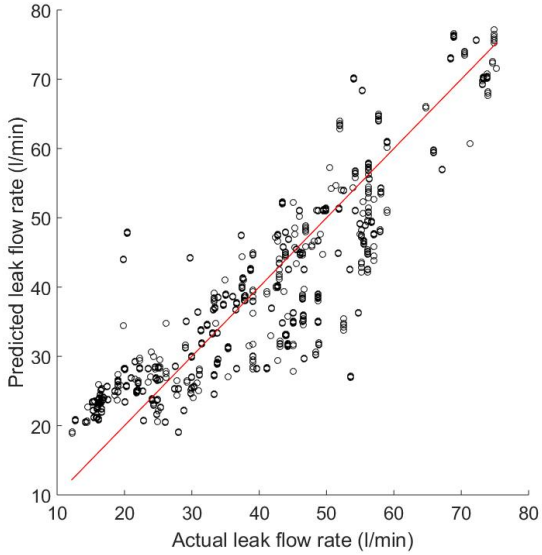


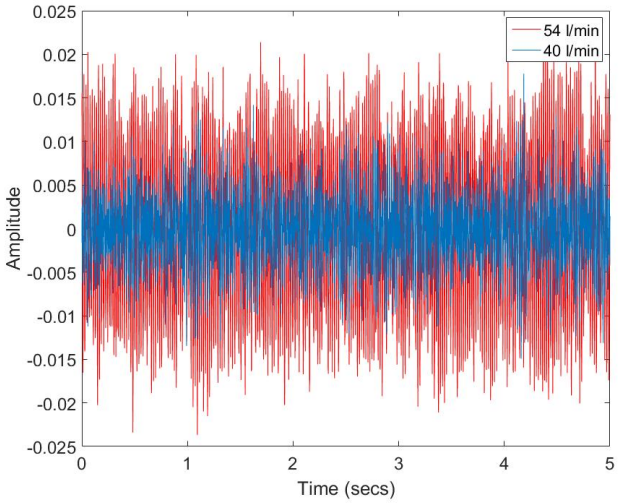
Figure 5.3: *Predicted vs actual (measured) leak flow rates using a regression model with 2nd order polynomial model. R^2= 0.79.*

5.3.2 RMS of IMFs as a single feature model

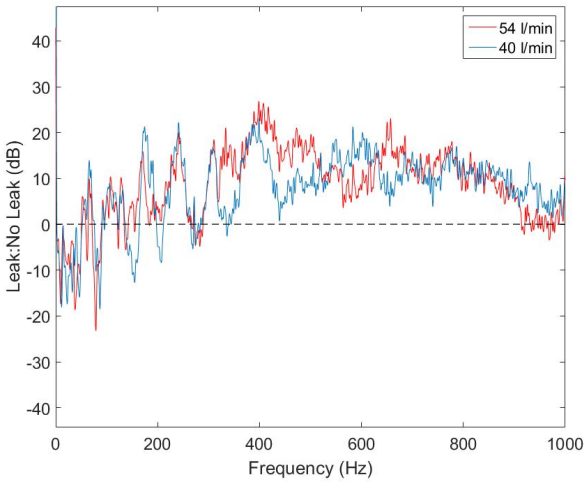
Table 5.2 showed that the RMS of both IMF1 and IMF2 may provide a useful feature in predicting leak flow rate due to the strong Pearson's Correlation Coefficient between leak flow rate and these 2 features (>0.8 for both features). Leak signals have been shown to vary in both time and frequency domain (Ahadi and Bakhtiar, 2010) and therefore features which describe both of these dimensions may provide useful features. Moreover, a study by Sun et al. (2016) utilised the signal decomposition method of EMD, showing that the RMS of individual IMFs can be used in order to identify the aperture of leaks from VAE signals. Therefore this parameter may be a useful in predicting leak flow rate. However, there are a number of weaknesses with the EMD calculation and the EEMD provides a more reliable method of decomposition (discussed in Chapter 2).

The process of EEMD breaks down signals in to individual IMFs, where each IMF represents differing frequency content of the signal. For demonstration purposes, 2 signals are demonstrated in Figure 5.4a and Figure 5.4b. These signals represent a 4.5 mm^2 round hole at 2 different leak flow rates of 40 l/min and 54 l/min. Both figures show an increase in signal amplitude and changes in frequency due a variation in leak flow rate. These two signals were subsequently decomposed via EEMD into individual IMFs. The corresponding first 6 IMFs in the time domain and the Fourier transforms of these IMFs are shown in Figure 5.5a and Figure 5.5b respectively. Each IMF has frequency components well below the pipe ring frequency (estimated to be in the region of 20 kHz for this pipe rig). The complexity of the leak signal is highlighted in that the separate IMFs are related to different frequency components of the leak signal, with a decrease in frequency as IMF number increased. This case of decreasing signal frequency as IMF number increases is typical of EEMD decomposition, as the sifting process removes higher frequency components first. Therefore, locally, an IMF will contain lower frequency oscillations than the previously extracted IMF. The EEMD decomposition identified signal in the region of 0-1000 Hz. The highest frequency components were identified in IMF1 and found to exist between 150 and 1000 Hz. However, these were distinctly low amplitude, with the highest amplitude signals of this band between 460-505 Hz,

but in general this IMF had low amplitude signals. The higher number IMFs were related to lower frequency oscillations, and in general those IMFs >IMF1 were of greater amplitude compared to IMF1. However, the highest IMF noted here (IMF6) represent very low frequency components and IMFs higher than this were regarded as background noise (IMF7 contains <28 Hz and in accordance with Figure 4.2 this is equivalent to background noise). The difference in signal amplitude within the different IMFs is also demonstrated in Table 5.4 whereby the higher flow rate leak had higher RMS values in IMFs 1-3 and IMF6. However, in the case of IMF4 and IMF5, the lower leak flow rate had a higher RMS value. The data presented here therefore suggests that a leak signal will have a different frequency and energy distribution through different IMFs depending on leak flow rate.



(a)

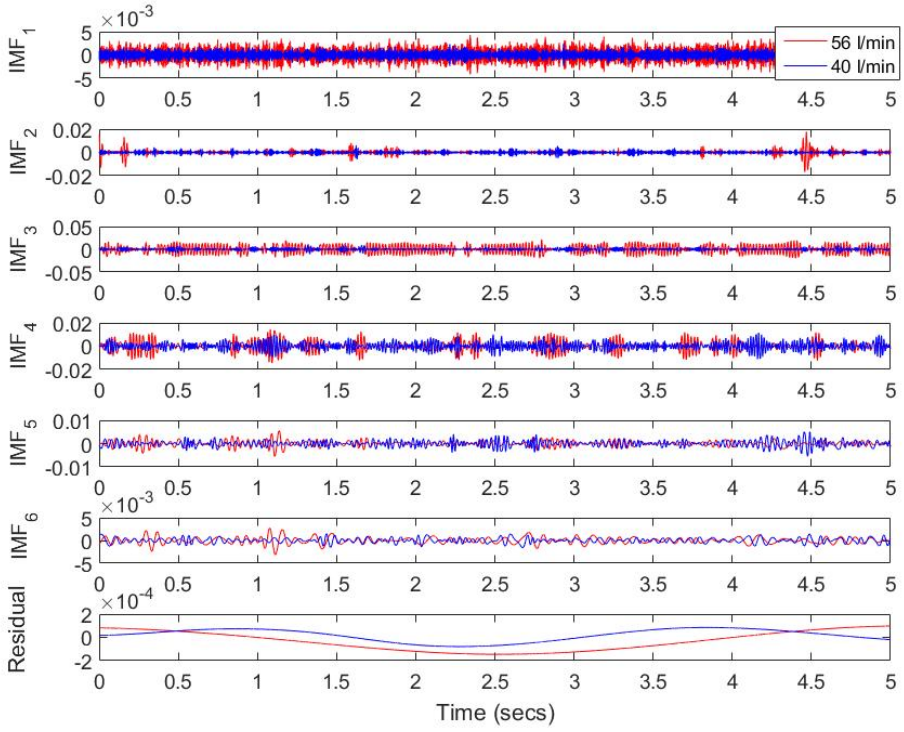


(b)

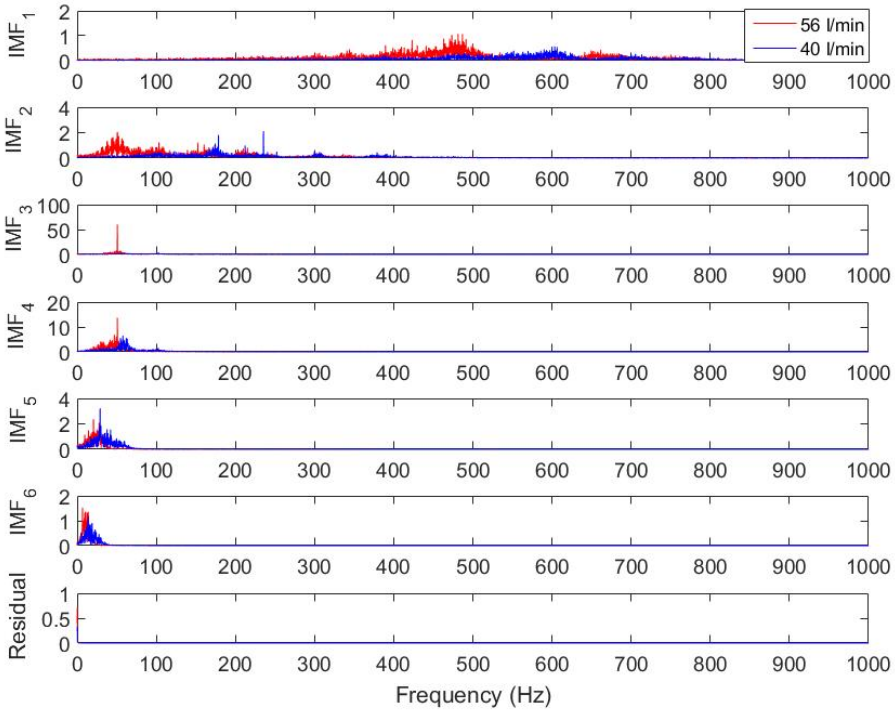
Figure 5.4: *(a) Representation of a high and low flow rate leak (a) time domain; and (b) ratio leak:no leak.*

Table 5.4: *RMS values of each IMF for low and high flow rate leaks. RMS is measured in mV.*

Leak flow rate	RMS IMF1	RMS IMF2	RMS IMF3	RMS IMF4	RMS IMF5	RMS IMF6
54 l/min	1.19 mV	1.81 mV	3.97 mV	1.16 mV	1.04 mV	0.70 mV
40 l/min	0.68 mV	1.00 mV	2.30 mV	3.14 mV	1.27 mV	0.53 mV



(a)



(b)

Figure 5.5: *Plots of IMFs following EEMD:(a) Time domain representation; and (b) FFT of each IMF. Comparison of a 40 l/min flow rate leak (blue) and a 56 l/min flow rate leak (red).*

The performance of the RMS of IMF1 as a single feature to predict leak flow rate is show in Table

5.5, which details the comparison of linear regression and non-linear regression models as a method of leak flow rate prediction using the RMS of IMF1 as a single feature input. The comparison demonstrates that the 3^{rd} order Polynomial model provides the best model for predictions, due to the fact that it achieved the lowest RMSE of 9.88 l/min. The curve fit to demonstrate the data points is shown in Figure 5.6, revealing a close fit of data to the 3^{rd} order Polynomial. The output of the best model is also shown in Figure 5.9, providing a training RMSE and test RMSE of 8.41 l/min and 9.88 l/min respectively. Again, there was no significant difference between the models and any model is likely to make a similar prediction of leak flow rate (due to the similar RMSE observed in Table 5.5). The RMS of IMF1 provided leak flow rate predictions with a higher RMSE than RMS of the whole signal, and therefore this parameter is less accurate as single feature model to predict leak flow rate compared to signal RMS.

Table 5.5: *Assessment of the RMS of IMF1 based model, comparisons of linear and non-linear kernels.*

Model	Training RMSE	Test RMSE	R^2	Equation (Q =)
linear	8.49	9.97	0.56	$26.43\text{mV}_{\text{RMS}} + 16.05$
Poly 2^{nd} order	8.42	9.89	0.57	$-4.59\text{mV}^2_{\text{RMS}} + 13.69\text{mV}_{\text{RMS}} + 11.54$
Poly 3^{rd} order	8.41	9.88	0.57	$2.67\text{mV}^2_{\text{RMS}} - 14.38\text{mV}^2_{\text{RMS}} + 46.87\text{mV}_{\text{RMS}} + 8.62$
Poly 4^{th}order	8.41	9.90	0.57	$-0.95\text{mV}^4_{\text{RMS}} - 2.17\text{mV}^3_{\text{RMS}} 6.13\text{mV}^2_{\text{RMS}} + 41.26\text{mV}_{\text{RMS}} + 9.77$
Logarithmic	8.67	10.18	0.54	$22.96\ln(\text{mV}_{\text{RMS}}) + 45.04$
Exponential	9.35	10.99	0.45	$19.66e^{0.69\text{mV}_{\text{RMS}}}$
Power	8.55	10.04	0.47	$41.99\text{mV}^{0.62}_{\text{RMS}}$

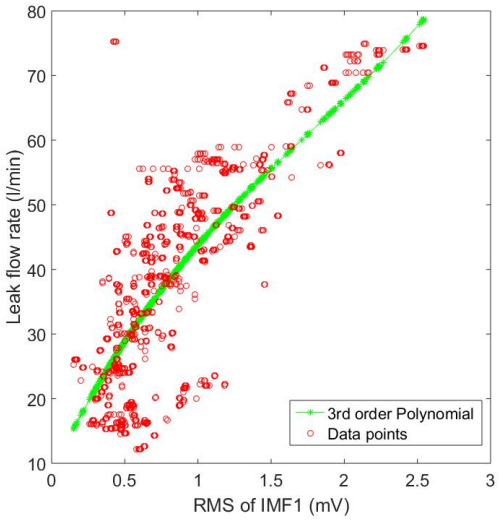


Figure 5.6: *Leak flow rate vs RMS of IMF1 for all leak area, shapes and backfill types. The data is fitted with a 3^{rd} order Polynomial curve. R^2= 0.57.*

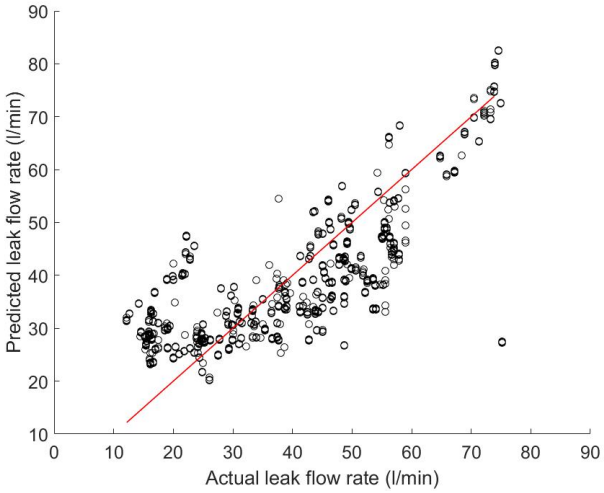


Figure 5.7: *Model predicted values vs actual values using RMS of IMF1 and a 3rd order Polynomial model. R^2= 0.57.*

The correlations shown in Table 5.2 also demonstrated that the RMS of IMF2 provided good correlations with leak flow rate, indicating that it may be a useful predictor in a model. The ability of this feature to predict leak flow rate is assessed in Table 5.6, where again, linear and non-linear models are compared. It was found that the model RMSE varied between 8.73 l/min and 10.43 l/min. The best model performance was using a 3rd order Polynomial which had a training RMSE of 8.73 l/min and test RMSE of 8.91 l/min, with an R^2 value of 0.68. The resultant fit of the model to the 3rd order Polynomial is shown in Figure 5.8. Generally, there is a good fit to the dataset, apart from at the lower leak flow rates (10-20 l/min), some outliers exist which may cause a reduction in model accuracy and associated R^2 values. The output of the model using a 3rd order Polynomial is displayed in Figure 5.9.

Table 5.6: *Assessment of the RMS of IMF2 based model as a single feature model to predict leak flow rate. Comparisons of linear and non-linear kernels.*

Model	Training RMSE	Test RMSE	R^2	Equation (Q =)
Linear	8.83	8.97	0.68	$19.21\mathrm{mV_{RMS}} + 13.05$
Poly 2nd order	8.77	8.92	0.67	$-1.59\mathrm{mV^2_{RMS}} + 24.24\mathrm{mV_{RMS}} + 10.15$
Poly 3rd order	8.77	8.91	0.68	$0.05\mathrm{mV^3_{RMS}} - 1.87\mathrm{mV^2_{RMS}} + 24.63\mathrm{mV_{RMS}} + 10.00$
Poly 4thorder	8.73	8.91	0.68	$-1.23\mathrm{mV^4_{RMS}} + 9.55\mathrm{mV^3_{RMS}} - 25.86\mathrm{mV^2_{RMS}} + 47.35\mathrm{mV_{RMS}} + 3.65$
Logarithmic	9.30	9.46	0.63	$21.58\ln(\mathrm{mV_{RMS}}) + 35.72$
Exponential	10.32	10.43	0.67	$17.63e^{0.53\mathrm{mV_{RMS}}}$
Power	8.89	9.05	0.63	$32.46\mathrm{mV^{0.62}_{RMS}}$

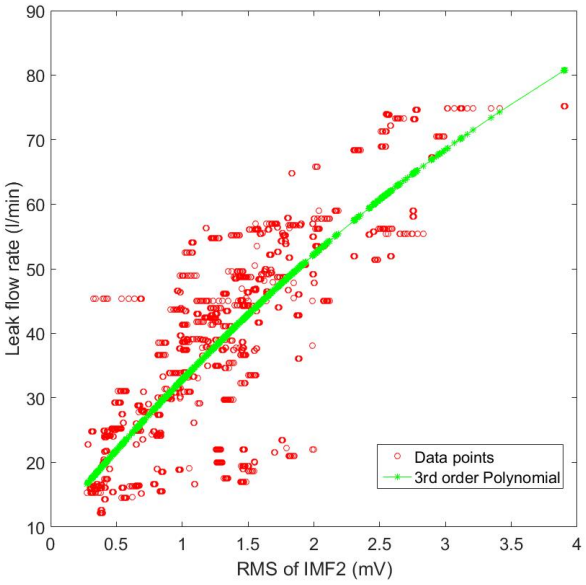


Figure 5.8: *The output of the IMF2 with a 3rd order Polynomial model. This model achieved an RMSE of 8.77 l/min and an R^2 of 0.68.*

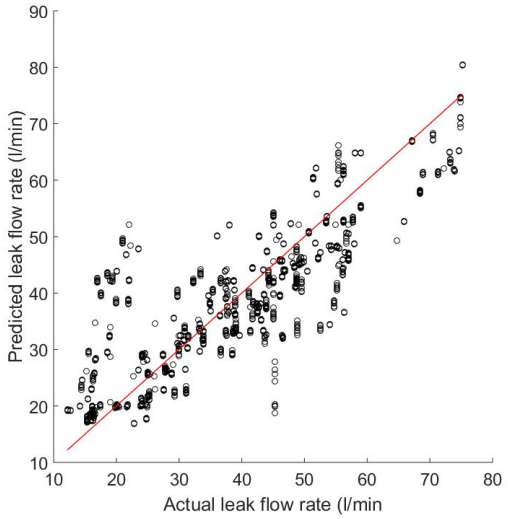


Figure 5.9: *Model predicted leak flow rates vs actual leak flow rates using a 3rd order Polynomial model. R^2= 0.68.*

The results demonstrated furthermore highlight that there can be alternate energy distribution in different IMFs depending on leak flow rate. However, the other IMFs did not have good correlations with leak flow rates and all had correlation coefficients of <0.44 (Table 5.2). Moreover, the other features calculated from EEMD which included the entropy of individual IMFs, were found to not correlate as well with leak flow rate (Table 5.2) and therefore single feature models were not constructed from these features.

5.3.3 Sensitivity of the single feature model

The analysis above identified that the best single feature model to predict leak flow rate was using the signal RMS. This was shown through the lowest training and test RMSE scores. Of interest is the effect of various variables on this model, questioning whether or not it is possible to improve on the current model results by isolating individual variables, such as the leak shape and backfill type (i.e. if the shape of a leak or backfill type was known prior to conducting training this may improve results). The following analysis re-trains and tests the RMS linear model with the leak shapes and backfill types known prior to training the model. Initially, leak shape is known to the model and the results of this are shown in Table 5.8. As the difference in model types (linear or non-linear) were negligible, a linear model was chosen for the following analysis due to its greater degree of simplicity over the alternate models. It was found that the models of electrofusion joints and longitudinal slits can be improved by training and testing the model when the shape is known prior to conducting the analysis. This is shown by a reduction in test RMSE to 3.09 l/min for longitudinal slits and 6.15 l/min for electrofusion joints from an original test RMSE of 7.57 l/min for all leak shapes. However, the RMSE of the round holes tended to increase for both training and test data compared to the global fit are shown in Figure 5.3, where the training RMSE and test RMSE increases from 6.29 l/min to 7.57 l/min respectively to 8.85 l/min and 8.85 l/min respectively. Further analysis of the specific shape-fit models shown in Figure 5.10 demonstrates that the round hole data presented in Figure 5.10a contains a number of outliers which may have negatively influenced the models performance.

Table 5.7: *Effect of isolating leak shapes on the linear model.*

Trained on	Training RMSE	Test RMSE	R^2	Equation (Q =)
Round Holes	8.85	8.87	0.72	$3.10mV_{RMS} + 18.89$
Longitudinal Slits	3.01	3.09	0.98	$3.67mV_{RMS} + 11.53$
Electrofusion Joints	6.09	6.15	0.75	$2.84mV_{RMS} + 24.02$

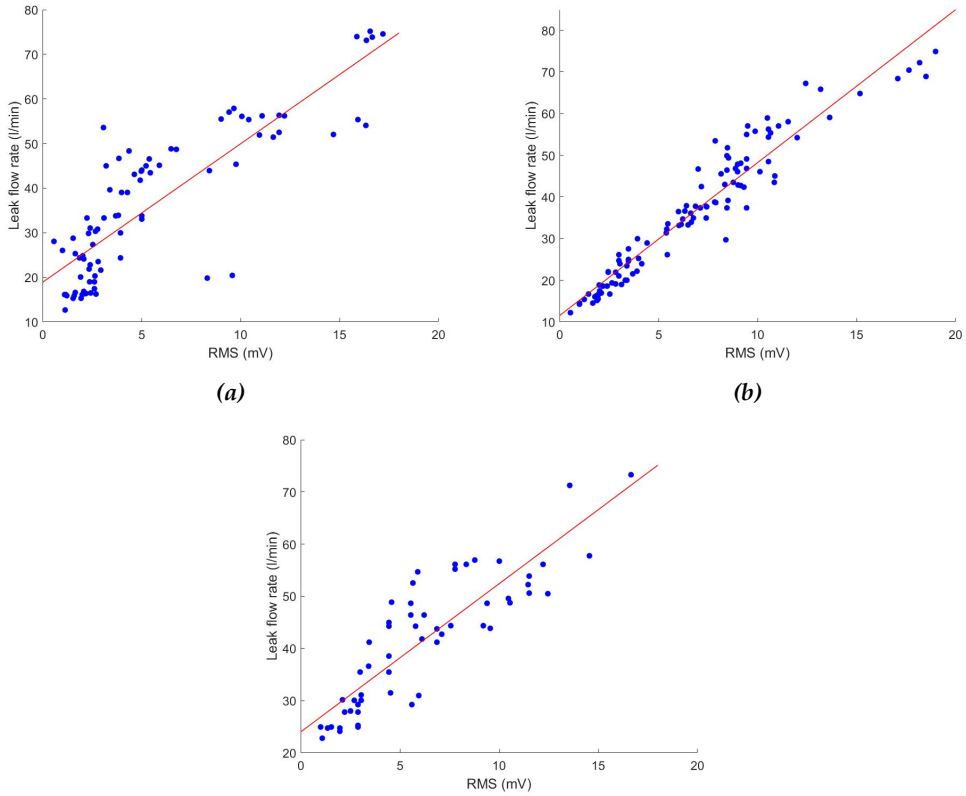


Figure 5.10: *Assessment of the effect of isolating individual shapes on the linear prediction model: (a) Round Holes (R^2= 0.72); (b) Longitudinal Slits (R^2= 0.98); and (c) Electrofusion Joints (R^2= 0.75).*

The model was also specifically trained and tested on different backfill types, isolating the effect of backfill on the model. The results for this are demonstrated in Table 5.8. All three backfill types resulted in a lower RMSE compared to that of the original data set and therefore should the backfill type be known prior to conducting training of the model, then the model accuracy can be improved.

Table 5.8: *Effect of isolating backfill type on the linear model.*

Trained on	Training RMSE	Test RMSE	R^2	Equation (Q =)
Gravel	6.60	6.68	0.77	$3.16mV_{RMS} + 16.59$
Submerged	6.62	6.64	0.77	$3.16mV_{RMS} + 18.44$
Geotextile	6.25	6.28	0.80	$3.24mV_{RMS} + 16.74$

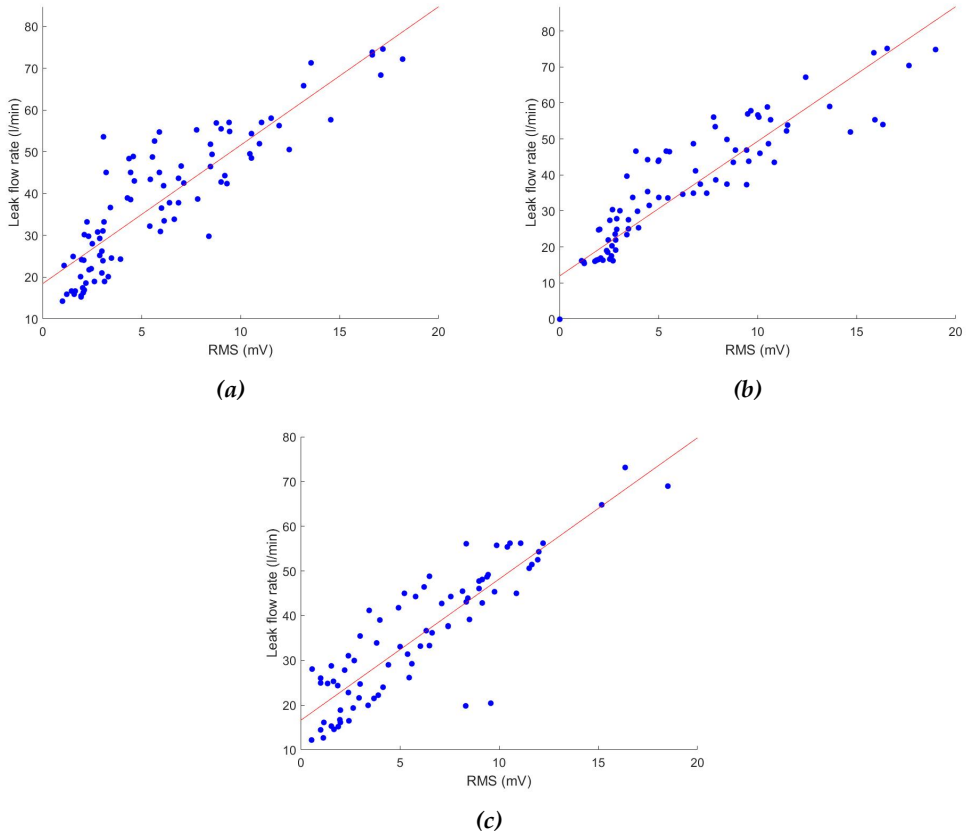


Figure 5.11: *Assessment of the effect of isolating individual backfill types on the linear prediction model: (a) Gravel (R^2= 0.77); (b) Submerged (R^2= 0.77); and (c) Geotextile Fabric (R^2= 0.80).*

5.3.4 Discussion of single feature models

The results presented so far in this chapter have investigated a number of features which were thought to have the potential to make estimates of leak flow rate in combination with linear and non-linear models. The analysis of the 22 features and how well they correlate with leak flow rate is given in Table 5.2. It was found that the signals which reflected changes in signal amplitude correlated highly with increases with leak flow rate. This is coherent with the findings in Chapter 5 where it was discussed in detail that higher leak flow rates increase leak amplitude. Therefore, it seems obvious that the signal RMS and RMS of IMF1 and IMF2 would be good leak flow rate predicting features as these reflect signal energy (Kaewwaewnoi et al., 2010). Frequency domain features which describe statistical characteristics of the signal (e.g. Skewness) were found not correlate well with leak flow rate and subsequent models were not built using these features. Of all the single feature models, it was found that the RMS of signals provides the best results. This is in agreement with both Chen et al. (2007), Kaewwaewnoi et al. (2010) and Kaewwaewnoi et al. (2007) who also showed increasing signal RMS with leak flow rate.

The sensitivity of RMS values as a single feature was investigated by isolating individual leak shapes and backfill types and creating seperate single feature models. The results suggest a sensitivity to leak shape and backfill, whereby a different model will be trained depending on the variable selected. This seemed especially prevalent in the case of leak shapes, whereby if the leak shape is known prior to training the model prediction accuracy can be greatly improved. However, this was only the case for the longitudinal slits and electrofusion joints. In the case of

the round holes, the model yielded higher RMSE when training only on round holes suggesting prediction was worst in this case. Whilst the model showed some sensitivity to changes in backfill type, it appears that this effect is much less in comparison to leak shape. However, the effect of backfill may become more important as sensors are moved further away from the leak and there is a greater effect of transmission parameters (backfill, pipe material etc.) on the leak signal. These results agree with Meland et al. (2011a) in that the calibrated models are only translatable to other variables when the model is trained and tested on all data (i.e. different models are produced when isolating individual variables). Unfortunately, this highlights a major limitation of the study as it was not possible to replicate all leak shapes and backfill types and therefore limits the applicability of this model to real WDS whereby there are many leak shapes, sizes and backfill types.

The results presented in this section suggests that the models based on single features would be less useful in real WDS as a wide variety of leak shapes and backfills can exist, amongst other variables. Indeed, every leak is unique, and no two leaks are likely to be the same. Moreover, parameters such as RMS are heavily reliant on the transmission istics of the pipe (i.e. the complex nature of a WDS can easily have an impact on the RMS values). Therefore, a model which combines a number of features together may provide the most fruitful of models as it is likely to be more robust to changes in other variables.

5.4 Multi-feature models

The RMS model and the RMS of IMF1 and IMF2 showed good correlations with leak flow rate and was found to provide good leak flow rate predictions through the development of a linear least squares model. However, single feature models by their nature are less robust, and the accuracy of these models are likely to reduce when there are more interacting variables that are common in real WDS. Moreover, features such as signal RMS change values depending on the distance from the leak and the presence of discontinuities in the pipeline. Therefore the presence of complexities in real world WDS means that these features alone are less likely to provide accurate results in a real systems. An optimal model for leak flow rate prediction is one that will learn and change depending on the characteristics of the pipeline system. Machine learning tools were discussed in Chapter 2 as one of the methods that may provide this. The following section assess whether the use of multiple feature models provide a better method of predicting leak flow rate compared to the single feature models already presented.

5.4.1 Data inputs

The only inputs in to the models were the 22 features (Tabl 5.1) for the leak flow rates studied, i.e. the models were not told anything else about the leak (such as it's flow rate, shape, backfill or area).

5.4.2 Multiple linear and non-linear regression

The first multi-feature model assessed was a multiple linear regression model which was created using all 22 features as inputs to the model. A plot of the output of the model (i.e. the predicted leak flow rates) compared with actual leak flow rates is provided in Figure 5.12. Note it now appears that there is a substantially larger data set within this plot which is due to the fact that this now contains

all the 22 features (rather than previous plots which only contain one feature). The multiple feature linear regression model was able able to predict leak flow rate with a training RMSE of 7.01 l/min and a test RMSE of 7.12 l/min ($R^2 = 0.79$). The model appeared to perform better at the lower leak flow rates, up to approximately 25 l/min, but then observed poor performance at leak flow rates >25 l/min. The greatest range of variation in predictions occurred at the mid-range of leak flow rates, which showed poor prediction accuracy, with the most inaccurate leak flow rate predictions at 55 l/min.

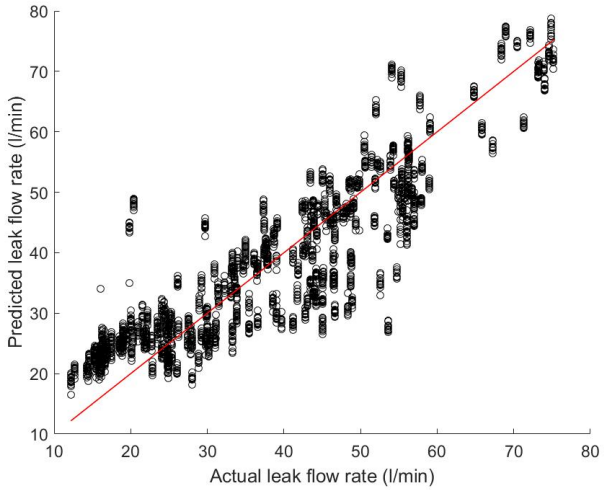
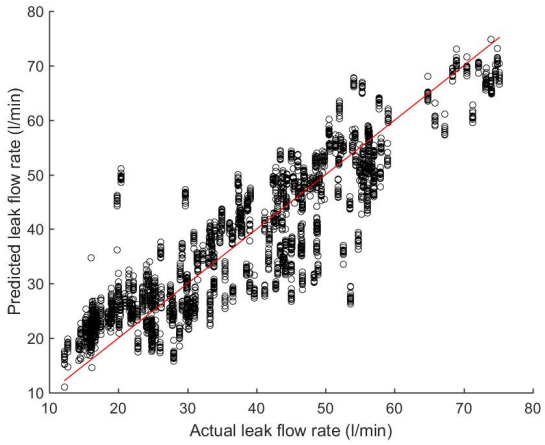
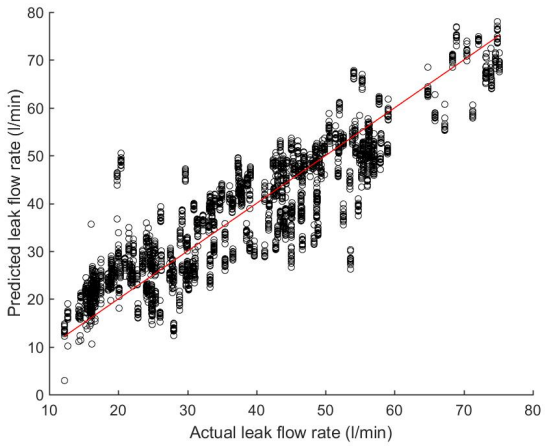


Figure 5.12: *Predicted and actual (measured) leak flow rates when using a multiple linear regression model. R^2= 0.81.*

As the single feature model analysis showed some slight improvements to model accuracy using the 2^{nd} order and 3^{rd} order Polynomial kernels, a multiple non-linear regression model was also carried out with the two aforementioned kernels. The outputs for both of these models are provided in Figure 5.13a and Figure 5.13b for the 2^{nd} order and 3^{rd} order Polynomial kernels respectively. It was found that the 2^{nd} order Polynomial model achieved a training RMSE of 6.99 l/min and a test RMSE of 7.03 l/min. The 3^{rd} order Polynomial kernel model obtained a slightly improved training RMSE of 6.95 l/min and a test RMSE of 6.98 l/min, an improvement from the linear model. Of particular interest is the improvement in the accuracy of leak flow rate estimates at the mid range leak flow rates when utilising a 3^{rd} order Polynomial kernel compared to a 2^{nd} order Polynomial kernel.



(a)

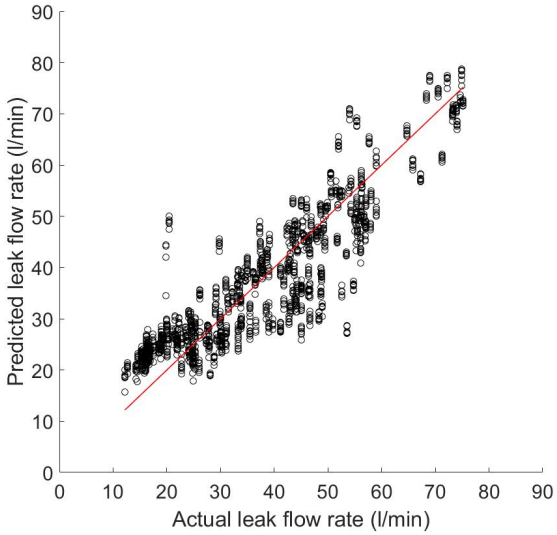


(b)

Figure 5.13: *Non-linear multiple regression model to predict leak flow rate: (a) using a 2^{nd} order Polynomial (R^2= 0.89); and (b) using a 3^{rd} order Polynomial (R^2= 0.91).*

5.4.3 Linear and non-linear Partial Least Squares Regression

The prediction of leak flow rate using linear PLSR and non-linear kernel based PLSR (KPLSR) is assessed here, following the work by Meland et al. (2011a). Non-linear Polynomial, exponential, power and logarithmic kernels are compared to a linear PLS model. It was found that the best results were achievable using a 3^{rd} order Polynomial, achieving a training RMSE of 7.01 l/min and test set RMSE of 7.25 l/min. The output for this is displayed in Figure 5.14a. The addition of a further order Polynomial made no difference to the RMSE. However, there was very negligible difference to the model output using KPLSR and the linear version (PLSR), whereby the linear PLSR had a training RMSE of 7.06 l/min and test RMSE of 7.31 l/min ($R^2 = 0.79$). The KPLSR model also showed increased accuracy compared with the single feature models. The KPLSR model tended to perform well at lower leak flow rates, but had more difficulty in making predictions at flow rates >30 l/min, with greatest prediction inaccuracy occurring at mid range leak flow rates (30 l/min to 60 l/min).



(a)

Figure 5.14: *A non-linear PLS model (3rd order Polynomial) to predict leak flow rate. R^2= 0.91.*

5.4.4 Least squares support vector machines

The figures presented above have assessed the ability of various multi-feature models in combination with a select number of features to predict leak flow rate. However, models such as LS-SVM may provide more powerful methods of supervised learning which may be able to recognise patterns in the data set. To assess the ability of the LS-SVM model, all of the 22 features were entered in to an LS-SVM model in order to assess it's ability to predict leak flow rate. The selection of the best kernel is one of the most important parameters in the successful operation of the LS-SVM model (Zhan et al., 2015), and all of the investigations shown above are in support of this, demonstrating that non-linear kernels provide the most fruitful results. Therefore, the LS-SVM model is trained using the non linear 'Radial Basis Function (RBF)' kernel, and has also been shown to give good results by other authors (e.g. Zhan et al. (2015)). The RBF kernel is expressed as (Zhan et al., 2015):

$$K(x, x_i) = exp(\frac{-||x - x_i||^2}{\sigma^2}) \tag{5.1}$$

where σ is the parameter of the kernel.

Although on a single feature model basis, some of the features may not be useful due to poor correlation with leak flow rate (Table 5.2), the features may perform well as part of a combination of features. However, using a number of features together may increase computational time, especially in cases where redundant features are being used. Redundant features can be removed using a method of feature redundancy. In this study, the forward search feature redundancy method (discussed in Chapter 2) was used in order to reduce the number of redundant features used in the model. Following the removal of redundant features, the LS-SVM model was able to predict leak flow rate with an RMSE of 4.63 l/min ($R^2 = 0.95$), a significant improvement compared to the models investigated previously. The output of the feature redundancy procedure on all shapes, areas and backfill types using LS-SVM is shown in Table 5.9 which demonstrates the addition of features to the model and its impact on model RMSE. It is shown that the model begins with the

RMS values as a feature (which was discussed above as having the lowest RMSE in a single feature model), and then gradually adds additional or removes features depending on whether or not it improves the model RMSE. The model output of the LS-SVM model is presented in Figure 5.15.

Table 5.9: *Feature selection and reduction in feature redundancy when predicting leak flow rate using LS-SVM. The best combination of features is identified by assessing the RMSE.*

Features used*	RMSE (l/min)	Total No. of features used
1	6.66	1
1, 6	6.14	2
1, 6, 23	5.50	3
1, 6, 23, 2	4.92	4
1, 6, 23, 2, 3	4.62	5
1, 6, 23, 2, 3, 5	4.97	6

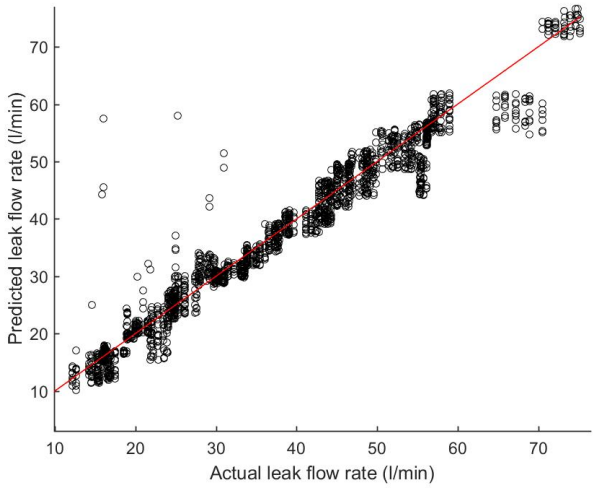


Figure 5.15: *Predicted and actual (measured) leak flow rates when using a LS-SVM model with RBF kernel. R^2= 0.97.*

Figure 5.16 shows a plot of the residuals using the LS-SVM model demonstrated in Figure 5.15 in order to understand the accuracy of the model at different leak flow rates. Incorporated in the plot of residuals are 2 bands demonstrating $\pm$ 10% of the actual leak flow rates. Of the flow rates studied, Figure 5.16 suggests that the model performed better at lower leak flow rates (apart from some outliers) and worst at the highest leak flow rates. The most accurate results came from the central band of leaks flow rates (between 35 l/min and 55 l/min) where predicted leak flow rate almost entirely remained within $\pm$ 10% of the actual leak flow rates. The model generally performed well, with a total of 81.99% of data falling within $\pm$ 10% of the actual leak flow rates suggesting a high prediction accuracy.

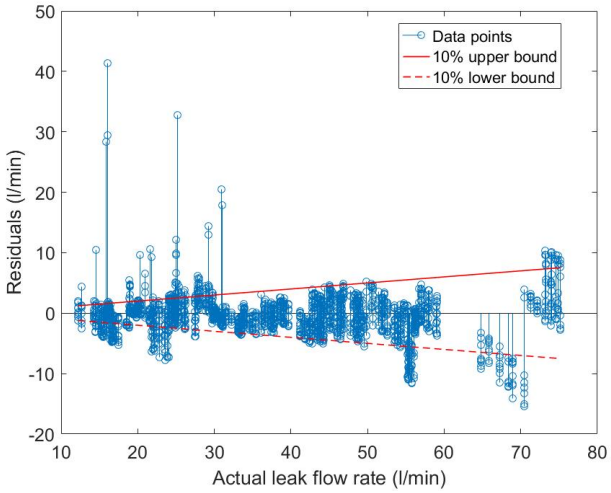
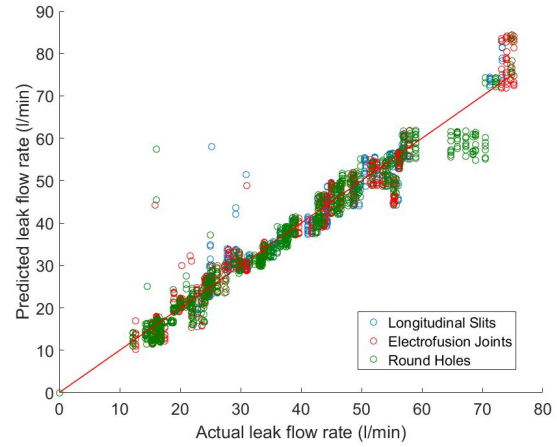
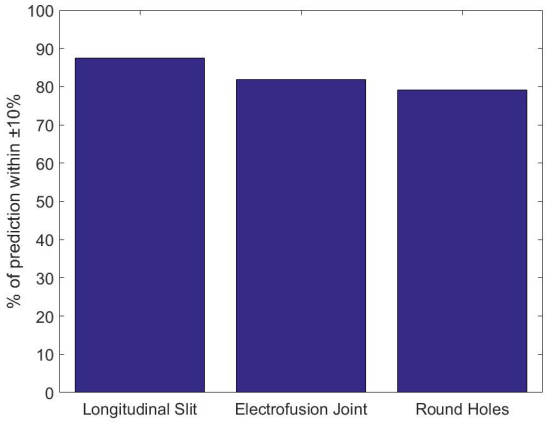


Figure 5.16: *Residuals of the LS-SVM model. ± 10% upper and lower bounds are also given.*

The LS-SVM model has provided the lowest RMSE in comparison with all the other models assessed in this chapter, suggesting that the LS-SVM model will provide the most accurate results in predicting leak flow rate. This was independent of changes to the backfill, leak shape and leak area. The sensitivity of the LS-SVM model to changes in leak shape and backfill type is now investigated in order to establish how robust the model is in response to changes in physical variables. Figure 5.17a colour codes the model in order to establish if there is an effect of leak shape on the models accuracy. It can be observed from Figure 5.17a that leak shape had a slight influence on the accuracy of the LS-SVM prediction results, however this appeared minimal. Moreover, the performance of the model in terms of it's prediction accuracy (± 10%) for different leak shapes is quantified in Figure 5.17b. When quantified in this way, it is appears that leak shape influences prediction accuracy, with the highest rate of accuracy being from predicting electrofusion joints with a prediction rate of 87.50%. Following this, the round holes predict to an accuracy of 81.85%. The worst results came from the longitudinal slits at 79.21%. However, generally the effect of shape appears to be minimal even when quantified in this way. Therefore, the accuracy of the model appears be negligibly influenced by leak shape, and therefore the model can be deemed robust to the changes in leak shape.



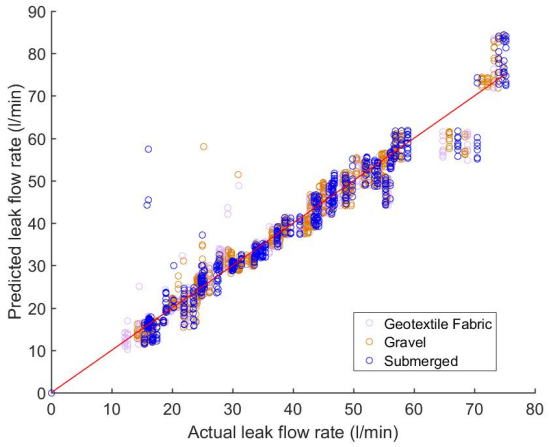
(a)



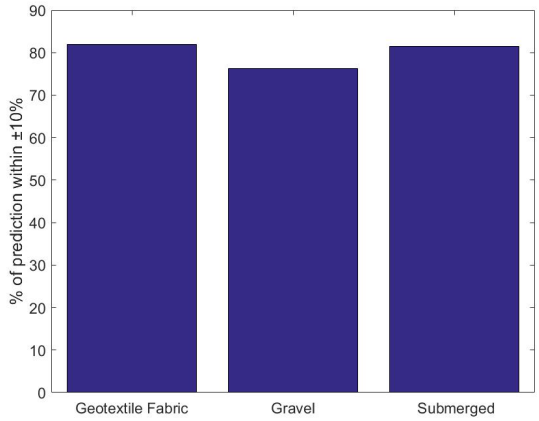
(b)

Figure 5.17: *Effect of leak shape on the LS-SVM model: (a) LS-SVM model output colour coded by shape (R^2= 0.97); and (b) accuracy of the model ± 10% of the actual leak flow rate.*

An investigation in to the influence of backfill on the performance of the LS-SVM model is demonstrated in Figure 5.18a and Figure 5.18b. The data shows that the greatest prediction error is from the gravel media, which accurately predicts 76.23% of data within ± 10% of the actual leak flow rates. The geotextile and submerged have a similar but much better prediction accuracy of 81.92% 81.49% respectively. Therefore it appears that it is still possible to achieve prediction accuracy of >76% independent of changes to backfill type.



(a)

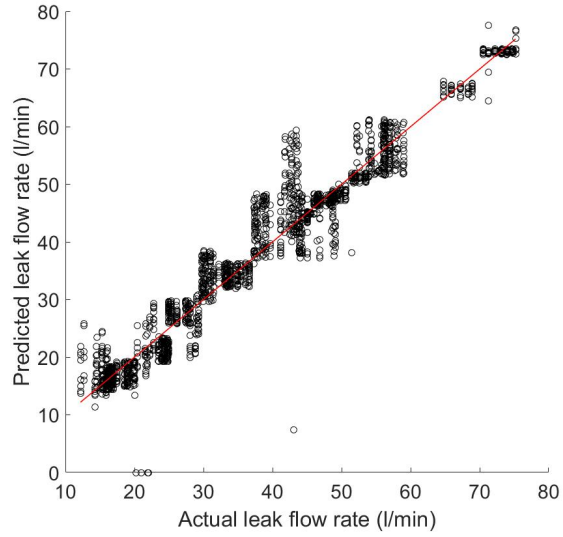


(b)

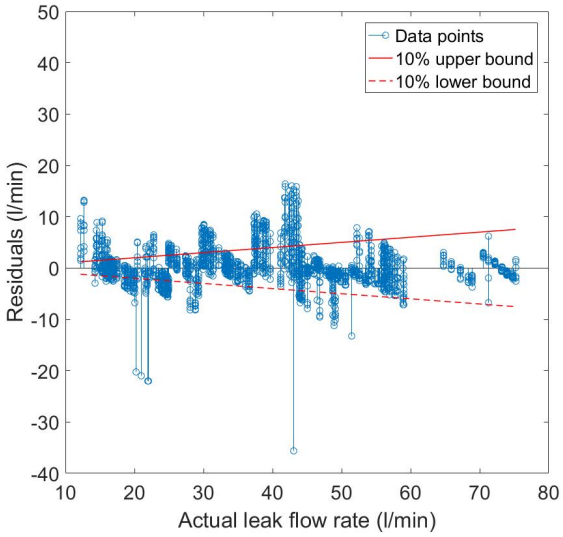
Figure 5.18: *Influence of backfill on the prediction results of the LS-SVM model: (a) model prediction vs actual leak flow rates (R^2= 0.97); and (b) effect of backfill on prediction results (within ± 10% of actual leak flow rate.)*

5.4.5 Comparison with hydrophone data

As mentioned in the Methods and Methodology chapter (Chapter 3), the leak signal was recorded simultaneously with both the hydrophone and accelerometer. As the hydrophone was placed slightly further away from the leak compared to the accelerometer shown in Figure 5.15 and Figure 5.17a, it is not possible to make a direct comparison between accelerometer and hydrophone signals. However, the evaluation of model accuracy is still worthwhile as hydrophones are receiving much attention in the literature due to their ability to provide improved correlations at reater distances compared with accelerometers (Gao et al., 2005). The efficacy of the hydrophone at Position P_9 (hydrophone position given in Table 3.1) in predicting leak flow rate using the LS-SVM model is shown in Figure 5.19a and Figure 5.19b. The performance of the hydrophone yielded a RMSE of 4.76 l/min (R^2= 0.94). The residuals plot in 5.19b demonstrates a high level of accuracy with the majority of samples within ± 10%. However, the hydrophone actually performed slightly worst than the accelerometer signal, although the differences in performance are possible more likely to be related to the fact the hydrophone is slightly further from the leak compared to the accelerometer.



(a)



(b)

Figure 5.19: *Prediction of leak flow rate using the hydrophone data: (a) actual vs predicted leak flow rates (R^2= 0.94); and (b) residual plots.*

5.4.6 Influence of sensor distance on model performance

So far, the work conducted has attempted to estimate the leak flow rate at the leak, in position P_1 and position P_2 for the hydrophones (see Table 3.1 for positions relative to the leak). However, it is unlikely that accelerometers will ever be placed this close to the leak when performing leak noise correlation on on real WDS, as accelerometers are normally placed on access points such as valves or fittings at distances away from the leak. Therefore the ability of the model to predict leak flow rate at distances away from the leak requires investigation in order to understand its ability to make leak flow rate predictions in real WDS.

Table 3.1 demonstrated the positions of the accelerometers as they are moved around the pipe rig. The ability of the model to predict leak flow rate at greater distances from the leak is now

assessed by running the model on accelerometer signals recorded in these positions, and the output of the LS-SVM model vs. the actual leak flow rates is given in Figure 5.20. The residual plot for each sensor position is also included in the figure. Figure 5.20a and it's residual in Figure 5.20b represents signals recorded at Position P_2 (2.25 m from the leak). Figure 5.20c and its residual in Figure 5.20c represent signals recorded in position P_5 (5 m from the leak). Figure 5.20e and its residual in Figure 5.20e represent signals recorded in position P_4 (8 m from the leak). These can be compared with Figure 5.15 which records leak signal at position P_1 (30 cm from the leak). The figures suggest that there is minimal difference between model accuracy between signals recorded at positions P_1 and P_2. The greatest differences come from comparing data acquired at position P_4 compared with data acquired at position P_1, although this does not seem to have a significant influence on the models accuracy. The signal recorded at position P_2 appeared to have slightly less accurate predictions at lower flow rates compared to position P_1, although this was relatively insignificant. Again, the model performed best at the middle leak flow rates at all sensor distances. The residual plots in all cases appear to show a higher degree of error at the lower and higher leak flow rates, whereas prediction accuracy was much improved at the mid-range leak flow rates. The residuals plots shown for each LS-SVM output plot show an interesting trend in the residuals. It appears that there is a similar pattern of error in prediction at all sensor distances, i.e. the error is similar at the same flow rates no matter what the distance. Therefore, it appears that the prediction error at these leak flow rates is independent of distance, and could be more to do with other variables such as leak shape or backfill.

Interestingly there was little difference between the LS-SVM model predictions at position P_5 (5 m from the leak, by the pipe bend) and position P_4 (8 m from the leak). This suggests that the bend had minimal effect on the performance of the model. This agrees with earlier results presented in this research that the bend had little impact on the leak signal (see Chapter 4). It also suggests that the effect of the flange plate between positions P_1 and P_2 had a greater influence on the model than the pipe bend.

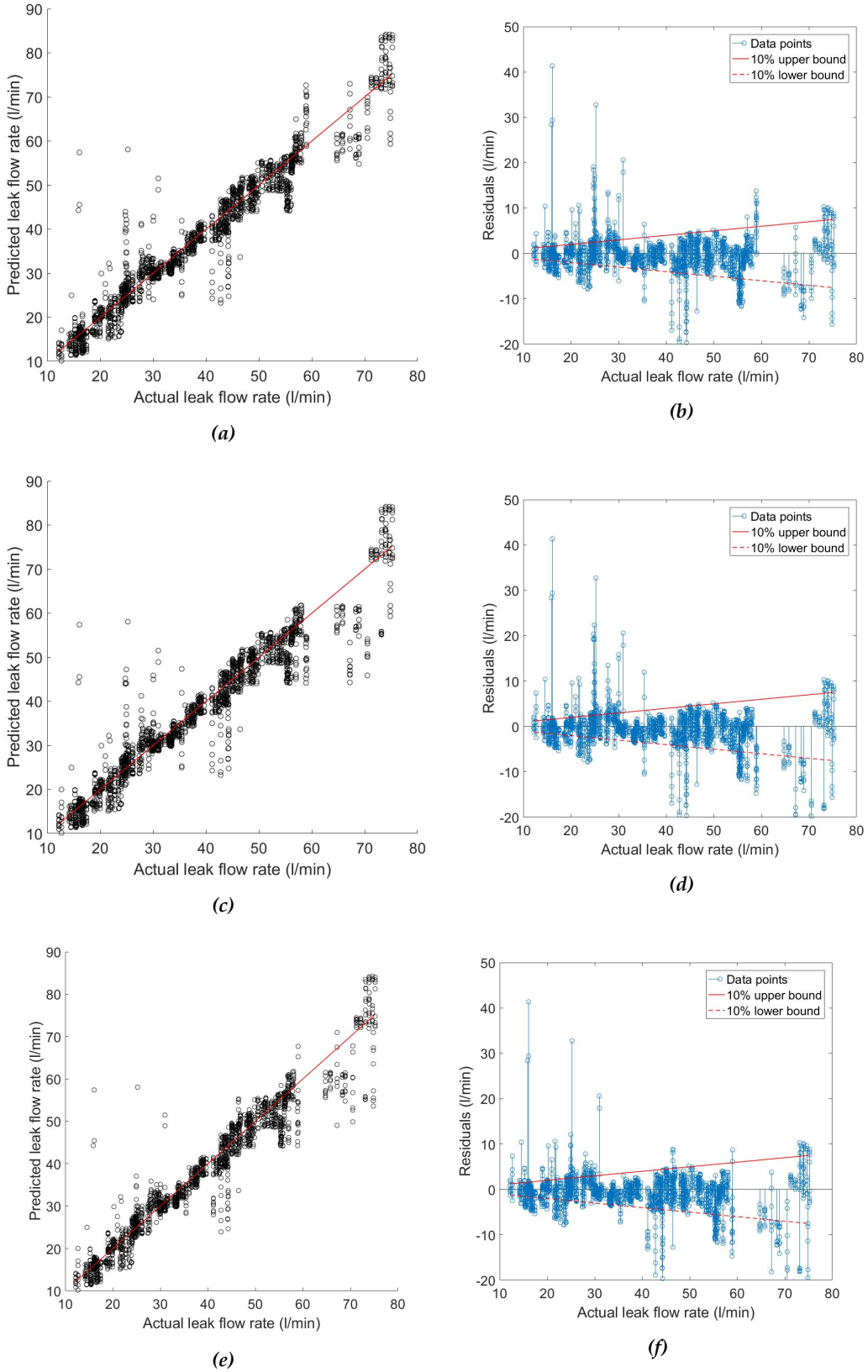


Figure 5.20: *Performance of the LS-SVM model in predicting leak flow rates at several positions away from the leak: (a) Position* P_1 *LS-SVM output* (R^2= *0.97); (b) Position* P_2 *residuals; (c) Position* P_5 *LS-SVM output* (R^2= *0.89); (c) Position* P_5 *LS-SVM residuals; (d) Position* P_4 *LS-SVM output* (R^2= *0.84); (e) Position* P_4 *residuals*

The recorded leak flow rates were from within a range of 12.90 - 75.19 l/min. Therefore, the simplest way to quantify the influence of sensor distance at the different leak flow rates is to divide them in to three separate categories, where: (1) leak flow rates $\leqslant 33.66 =$ 'low leak flow rate'; (2) >33.66 and <54.42 = 'medium leak flow rate'; and (3) $\geqslant 54.42 =$ 'high leak flow rate'. These categories where chosen to reflect the range of leak flow rates evaluated in this study. In real WDS, leak flow rates can be wide ranging depending on a number of factors, specifically leak area and system pressure. Therefore, the leak flow rate categories chosen in this study do not represent categories of leak flow rates in real WDS. Overall performance of the model in terms of its accuracy within $\pm$ 10% of actual leak flow rate, within these three bands is shown in Figure 5.21 (essentially, this is a count of the number of samples that are predicted to within $\pm$ 10% of the actual leak flow rate). The accuracy of leak flow rate prediction as a function of position is therefore slightly easier to quantify using this Figure. The figure shows that at all sensor positions, the mid-range leak flow rates always have the best prediction accuracy. The worst predictions overall are from higher leak flow rates. The lowest leak flow rates have a fairly consistent rate of prediction accuracy at approximately 73%. There does not appear to be a significant influence of transmission distance on the prediction results. However, prediction accuracy is slightly higher at position P_1 compared to the other sensor positions, and this seems to be the case at all leak flow rates. Predictions at position P_4 actually appear slightly better than at P_5, even though this is an extra 3 m in transmission distance. This seems to occur at the low and high leak flow rates, but at the mid-range leak flow rate there does not seem to be any changes in prediction accuracy.

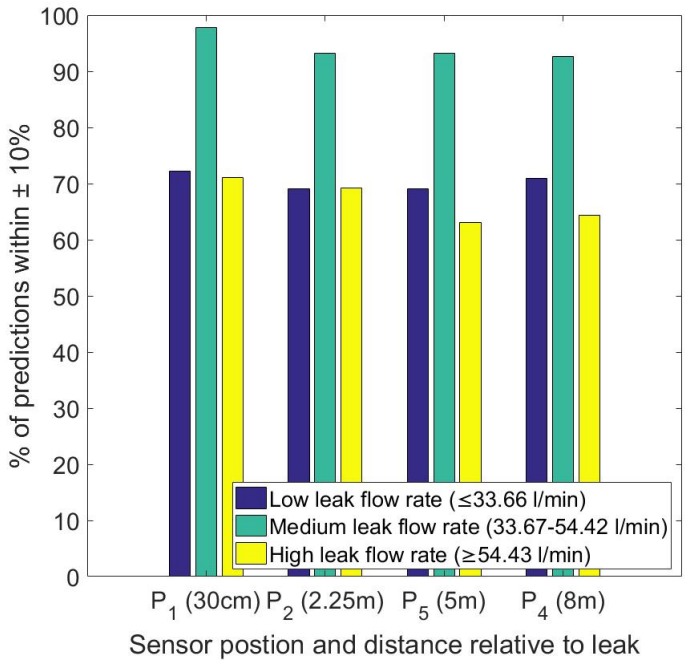


Figure 5.21: *Accuracy of the LS-SVM model at various sensor positions on the test rig. The results are given as the % of correct predictions within $\pm$ 10% of the actual (measured) leak flow rates). Distances from the leak are given and 'P$_1$', 'P$_2$', 'P$_3$' relate to positions given in meters and in Table 3.1.*

5.4.7 Multi-feature model discussion

The 22 features derived from the model were entered in to various multi-feature models in order to predict leak flow rate. The multi-feature models studied all led to decreases in model RMSE compared to the single feature models, suggesting increased model accuracy when using multi-feature models. The results presented herein are therefore coherent with similar findings shown by Meland et al. (2011a) who observed best leak flow rate predictions from gas pipes using multi-feature models, in which they found KPLSR to be the best model. However, the study presented herein found negligible improvements in RMSE when comparing the MLR, KMLR, PLSR and KPLSR models. Therefore, in terms of multi-feature models, it is probably not worth the extra computational cost in running models such as KPLSR and thus a simple MLR would provide fruitful results at the lowest computational cost.

This study reviewed an extensive quantity of literature (Chapter 2) and identified that the LS-SVM model may provide more accurate results, without jeopardising computation cost due to quick run times. Indeed, the LS-SVM model provided the most accurate results in predicting leak flow rate with the lowest RMSE of all the models investigated (single and multi-feature). The predictions with the LS-SVM were also shown to be robust to changing leak shapes and backfill types. However, it was found that this would be at a much greater computational cost compared to the MLR model (with LS-SVM models taking a greater time to train), and therefore the user needs to consider the trade-off between prediction accuracy and computational cost.

Analysis of the performance of the LS-SVM model at different leak flow rates showed that the LS-SVM model provided best leak flow prediction at the mid-range leak flow rates (Figure 5.15) with lower standard deviations and higher prediction accuracy. The model was able to predict mid-range leak flow rates to higher accuracy than leak flow rates at the extremities, as shown by the plot in Figure 5.16. This is surprising as leaks with larger flow rates are normally easier to identify (Humphrey et al., 2012) as signal amplitude increases with leak flow rate (Kaewwaewnoi et al., 2010). The model also performed better as the number of features were added, up to a total of 5 features, following this there was no significant increase in model performance as further features were added. This is further shown by the fact that the reduction in the LS-SVM model RMSE is large when increasing the number of features used from 1 to 5. The addition of further features did not lead to any increase in model RMSE, demonstrating the importance of using method to remove feature redundancy. This information suggests that not all 22 features are required and therefore these can be removed from the model as they will only introduce a greater degree of complexity and increased computational cost with no additional value to the model.

Consistent with the single feature models, it was found that features relating to signal amplitude were highly useful for predicting leak flow rate, whereby the RMS of IMF1, IMF2 were identified by the forward search algorithm as providing the optimal combination of features (Table 5.9). Signal parameters that describe leak amplitude are likely to be a useful parameter as amplitude has been shown to increase with leak flow rate by numerous authors (Chen et al., 2007; Kaewwaewnoi et al., 2010) and is coherent with the findings in Chapter 5. Therefore, it is unsurprising that these features were important to the model. Interestingly, the model also favoured features following signal decomposition via EEMD. The FFT of each IMF has demonstrated that those IMFs >IMF3 represent the extremity of the signals lowest frequencies and represent frequencies relating to the background noise (<28 Hz) identified by comparing the leak and no leak spectrum's (Figure

5.5b). Leak signals are inherently composed of leak signal and background noise (Gao et al., 2004) and the use of EEMD in this case can help to separate the leak signal from noise. As these higher numbered IMFs represent signals unrelated to te leak, the model prefers to predict leak flow rate based on the RMS of parts of the signal relating solely to the leak signal and excludes other noise (utilising IMFs 1, 2 and 3).

The model therefore appears to also have a strong preference to using time-frequency features produced by EEMD. The reasons for this are possibly tow-fold: (1) EEMD is breaking the signal in to time-frequency components and as the leak signal differs in time and frequency domain (Ahadi and Bakhtiar, 2010), the model is able to derive information from the first 3 IMFs of the signal; and (2) the subsequent calculation of RMS represented increases in signal amplitude for each IMF as leak flow rate increased and therefore it was possible to detect changes in amplitude as leak flow rate increased. It was found that when using EEMD the level of decomposition in each IMF changed depending on leak flow rate. This was shown explicably in Figure 5.5 whereby two signals of different leak flow rates are shown and the composition of signal within each IMF differs depending on the leak flow rate. As EEMD is a data-driven approach and is fully adaptable, the decomposition acts like a dyadic filter bank (Flandrin et al., 2004), whereby the process can be interpreted as a multitude of overlapping filters. The way in which the filter banks overlap depends largely on the signal and the extraction of the previous IMF. This approach means that filter banks depend on leak flow rates and therefore information will differ depending on this.

The decomposition of the leak signal in to features describing the time-frequency domain will become less useful and effective in predicting leak flow rates when the sensor is positioned further away from the leak. As sensors are positioned further away from the leak, there is a loss of higher frequency components and a general reduction in signal amplitude due to signal attenuation on the pipe wall and radiation in to the surrounding media (Almeida et al., 2014a). Thus the leak has a low signal to noise ratio (Humphrey et al., 2012) and is difficult to distinguish from the background noise. Therefore the features which are highly valued by this model will have reduced efficacy as these are only present in this study as the sensor was positioned close to the leak. It may then be more difficult to quantify leak flow rate using these features at greater distances from the leak and it may be more valuable to use other features which are focused around spectral shape as in reality it would be difficult to position a sensor at such close proximity to the leak. This highlights a major limitation of this research. However, it was not possible to place sensors at long distances away from the leak due to space restrictions within the laboratory and therefore this effect could not be investigated.

In addition, the LS-SVM model included the parameter crest factor which indicates the peaked-ness of the waveforms. Whilst there is no data to suggest that crest factor increased with leak flow rate (a low Pearson's Correlation Coefficient was found between the leak flow rate and crest factor), the LS-SVM models still found this feature to be useful as part of a combination of other features. Moreover, there was no model preference for such features that rely on the frequency domain characteristics, suggesting that frequency domain characteristics do not make useful features for leak flow rate prediction. The results presented herein therefore differ from that of Meland et al. (2011a) who suggested that frequency inputs where the most useful features.

The model was also trained with hydrophone data (Figure 5.19), producing an RMSE of 4.46 l/min, which is a slight improvement in model results compared to the accelerometer data which

achieved an RMSE of 4.63 l/min. This agrees with current understanding that hydrophones usually offer better performance for leak detection when compared with accelerometers (Almeida et al., 2014a; Gao et al., 2005). The reasons for the slight differences in model performance when using hydrophone data may be due to the influence of a smaller effective bandwidth and higher signal coherence (Almeida et al., 2014a) when using hydrophones. Also, as the accelerometer is placed on the pipe wall, it may be more susceptible to changes in ground conditions whereas the hydrophone will not be, as the ground conditions have been shown to have a strong influence on the leak signal (Muggleton and Brennan, 2004). Whilst efforts were made to ensure that the ground conditions were standardised, the replacement of the test sections resulted in the excavation of the pipe from the gravel backfill. Slight changes to the test conditions are possible through changes in loading, soil hydraulics and flow resistance which can have an influence on leakage dynamics (Fox et al., 2016). As the accelerometer is in contact with the pipe wall, the effect of ground conditions may be more paramount and therefore may influence the recorded signal. However, the hydrophone will likely be more susceptible to changes in pipe flow regime and flow turbulence. Moreover, whilst the hydrophone offered slightly improved performance, the difference was only negligible. Hydrophones are not favoured by water companies due to potential of contamination, resulting in risks to water quality (Pal, 2008) and therefore the results here suggest that the accelerometer should continue to be used due to the fact that there are only negligible differences in model performance when compared with the hydrophone.

5.5 Chapter summary

The research presented in this chapter has shown that with some additional signal processing, there is sufficient information contained in the leak signal in order to predict the leak flow rate without prior knowledge of leak area, leak shape or backfill type. This was initially achieved by deriving a specific set of features from the recorded leak signal and tested for their suitability. The RMS and RMS of IMF1 and IMF2 of the leak signal appeared to be a useful parameter for leak flow rate prediction. It was found that based on single features, simple regression models with non-linear kernels were able to predict leak flow rate. However, these single feature models alone where found to not produce robust leak flow rate predictions with high RMSE. Morevoer, due to the complexities of real WDS, a single feature alone as a leak flow rate prediction model is unlikely to be suitable. Therefore, a method which is adaptable and learns with more data is required.

It was proposed that the inclusion of a number of features and using a model which utilises a multitude of features may provide lower model RMSE. 22 features derived from the signal and were entered in to several multi-feature model types and compared. Of the models studied, the LS-SVM model provided the best RMSE and was able to predict leak flow rate despite changes to backfill, leak shape and leak area. Moreover, the LS-SVM model which utilised a combination of a number of features, was able to predict leak flow rate to greater accuracy than all of the other single feature and multi-feature models. The knowledge gained from this study, and the proposed technique, is important to leakage management as it will allow the leakage manager to prioritise repair strategies, reducing the volume of water lost and making the process of active leakage control more efficient.

Chapter 6

Leak area and leak shape prediction

6.1 Introduction

Chapter 4 highlighted the effect of leak area and leak shape on the leak signal in laboratory conditions with well controlled experimental methods. It was found that, when studying leaks of different shapes and sizes at similar leak flow rates, leak area had a negligible effect on the leak signal when observing frequency domain signals. The study also showed that leak shape was found to be a significant factor in characterising the leak signal. It was also suggested in Chapter 2 that some shapes have both a time and pressure dependent leak area and therefore a method which could classify the leak area and leak shape from the VAE signal would be useful for prioritising leak repair activities. The aim of this chapter is to derive a method that can predict both leak area and leak shape using the leak signal. This chapter is therefore divided in to two parts: Chapter 6a dealing with leak area prediction, and Chapter 6b dealing with leak shape prediction.

Whilst it is difficult to predict the area and shape of a leak by analysing the frequency spectrum alone (shown in Chapter 4), other features that describe the signal could provide more fruitful results. The same set of signal features which were derived in Chapter 5 are also applied here to develop a leak area and leak shape prediction model. However, whilst Chapter 5 was more suited to predictive models, this Chapter is more suited to classification based models.

Unlike in Chapter 5 which considered several leak flow rates between 12 and 72 l/min, this section only considers 5 leak flow rates, measuring approximately: 39-40 l/min; 44-45 l/min; 47-48 l/min; 49-51 l/min; and 56-57 l/min. Only 5 leak flow rates were selected as these are the leak flow rates which are standardised across all leak areas, leak shapes and backfill types. Therefore, the influence of leak flow rate on the signal is negated - if this is not done then the model is more likely classifying different leak flow rates rather than leak area or leak shape. This approach was taken because there is high value in being able to predict leak area and leak shape independent of leak flow rate. Leak flow rates were kept the same across all leak shapes and area sizes by controlling system pressure with the downstream gate valve. Each leak shape is also assessed on 3 backfill types at the 5 different leak flow rates. Again, each test was repeated 20 times.

6.2 (a) Leak area prediction

6.2.1 Introduction to leak area prediction

One of the key objectives of this research was to derive a method that can predict leak area solely using the VAE signal. A similar approach as Chapter 5 is required here whereby a model can be used to make predictions of leak area. This section therefore assesses some features which can be helpful to predict leak area, with a particular emphasis on the features which were identified to be useful by the LS-SVM model in Chapter 5. It is hoped that similar features can also be utilised in order to provide good estimates of leak area.

6.2.2 Feature assessment

The observations made in Chapter 5 revealed that the leak area appeared to have a negligible effect on the leak signal. This was shown in the frequency domain by taking the ratio of leak:no leak. However, due to negligible differences in signal with leaks of the same shape and leak flow rate but different area, the frequency domain representation offered little way of identifying the area of the leak. It may be possible that other features can describe the signal which could lead to a method of leak area prediction.

Chapter 4 also highlighted that there was a strong influence of leak area and leak shape on signal RMS level. However, it was difficult to classify leak area and leak shape by observing trends in this value alone. The investigation in Chapter 5 also revealed that the RMS of IMF1 and IMF2 were useful features for leak flow rate prediction. Moreover, they were found to be useful features in predicting leak flow rate using the LS-SVM model. Quite possibly, these features may help to classify leak area. Accelerometer signals recorded from position P_1 with the same leak shape, leak flow rate but different leak area can now be broken down via EEMD. The results of the EEMD breakdown of round hole leaks measuring 4.5 mm, 5.5 mm and 6.5 mm at a leak flow rate of approximately 40 l/min are presented in Figure 6.1 (these have been shown for example purposes). Again, it is important to note that these leaks are the same shape and same approximate leak flow rate, but are of a different leak area. It can observed that in terms of the frequency bands provided by EEMD breakdown, there are some slight difference between leak area and frequency within IMF1. However, there are more observable differences in frequency and amplitude of IMF2, IMF3 and IMF4. The RMS of IMF1 and IMF2 were shown in Chapter 5 as useful features for leak flow rate prediction and therefore the associated RMS of IMF1 and IMF2 is given in Table 6.1 for the 3 example signal shown in Figure 6.1. It is demonstrated that there are differences in RMS within both IMF's depending on leak area despite the approximate leak flow rate being the same for each leak area. Therefore, it is possible that the RMS of each IMF can contain enough information to distinguish between signals of different leak area. The results presented here appears to indicate that the process of EEMD can show differences in the signal based on changes to leak area. This method therefore highlights a way of distinguishing between leak area independent of leak flow rate, which the frequency domain calculation could not. This is therefore of high value to this study as it has managed to derive a feature which can distinguish between leak areas independent of leak shape and leak flow rate.

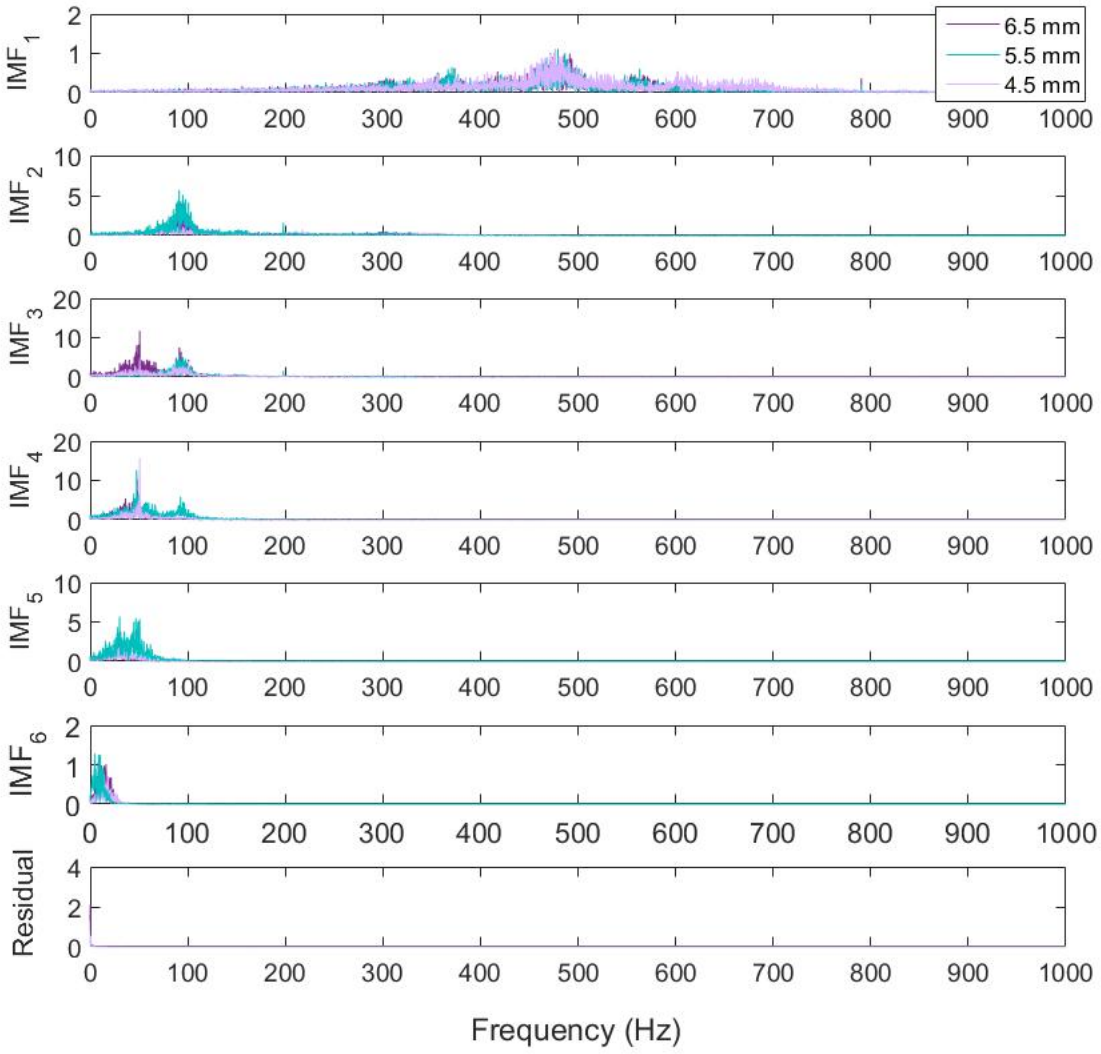


Figure 6.1: *Frequency domain breakdown of signal via EEMD. Round hole leaks measuring diameter 4.5 mm, 5.5 mm and 6.5 mm round hole at approximately 40 l/min are presented.*

The individual breakdown of the RMS of IMF1 and IMF2 for each leak flow rate are given in Figure 6.2 and Figure 6.3 respectively. The differences in leak area are also highlighted. It is shown that differences in leak RMS levels are found for different leak areas. However, it is difficult to establish whether the changes in RMS level is due to the increased leak flow rate or changes to leak area as leak flow rate increases for leak area when the system pressure is kept consistent. In order to establish the full effect of leak area on the RMS of IMF1 and IMF2, the downstream gate valve was controlled in order to ensure that the leak flow rates remained the same for different leak areas. Table 6.1 demonstrates the RMS of IMF1 IMF2 for different leak areas (3.5 mm to 6.5 mm) whereby the the same leak flow rate was achieved. It was found that the the RMS of IMF1 and IMF2 does vary slightly between changes in leak area, albeit this is negligible. Therefore, these results suggest that RMS does not change significantly between leak are and therefore it may not make a useful parameter in classifying leak area.

Table 6.1: *RMS of IMF1 and IMF2 for leaks presented in the above figures, buried in gravel backfill.*

Leak diameter	Leak flow rate (l/min)	RMS IMF1 (mV)	RMS IMF2 (mV)
4.5 mm	40	0.73	1.01
5.5 mm	40	0.68	1.41
6.5 mm	40	0.78	1.85

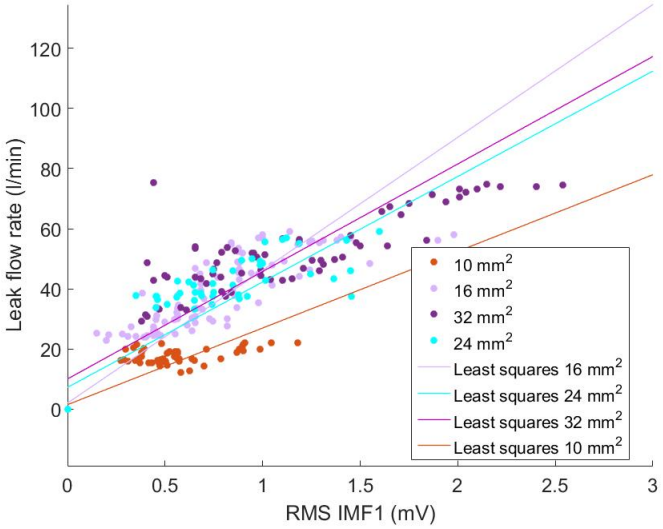


Figure 6.2: *RMS of IMF1 of all leak flow rates with data set divided depending on leak area, including area sizes of 10 mm^2 (R^2= 0.75), 16 mm^2 (R^2= 0.93), 24 mm^2 (R^2= 0.90), 32 mm^2 (R^2= 0.91).*

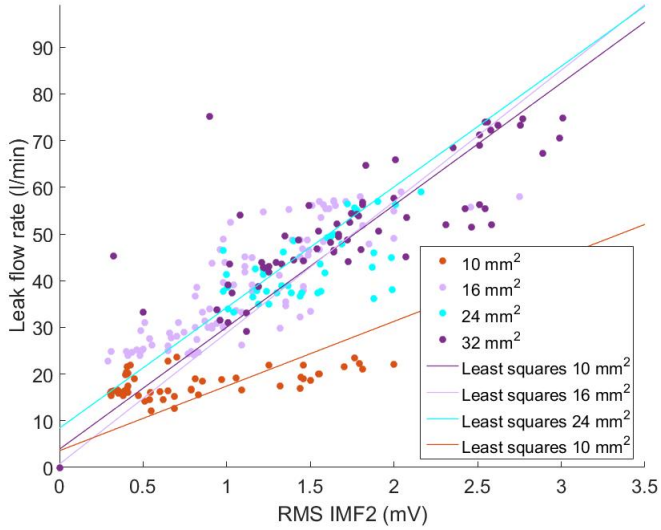


Figure 6.3: *RMS of IMF2 of all leak flow rates with data set divided depending on leak area, including area sizes of 10 mm^2 (R^2= 0.75), 16 mm^2 (R^2= 0.95), 24 mm^2 (R^2= 0.98), 32 mm^2 (R^2= 0.72).*

6.2.3 Prediction of leak area model performance

One of the objectives of this study is to develop a tool which predicts leak area from the leak VAE signal. Chapter 5 highlighted the fact that multiple-feature models continuously provide better performance in predicting leak flow rate. Here, categorical based multiple feature models are applied in order to predict leak area in this study. The models include Logistic Regression (LR) and Random Forest (RF), representing linear and non-linear models respectively.

6.2.3.1 Model inputs

Only the 22 features (detailed in Table 5.1) derived from the signal were used as inputs in to the model, i.e. the model was not told what the leak shape, leak flow rate or backfill the leak was present in. Initially, a test of linearity was conducted whereby all shapes were included in a model but only under gravel media. Following this, all data was added to the model, that being the 22 features from all leak shapes, leak areas, backfill types and leak flow rates. A simple process flow diagram of the approach to building the leak area prediction model is presented in Figure 6.4.

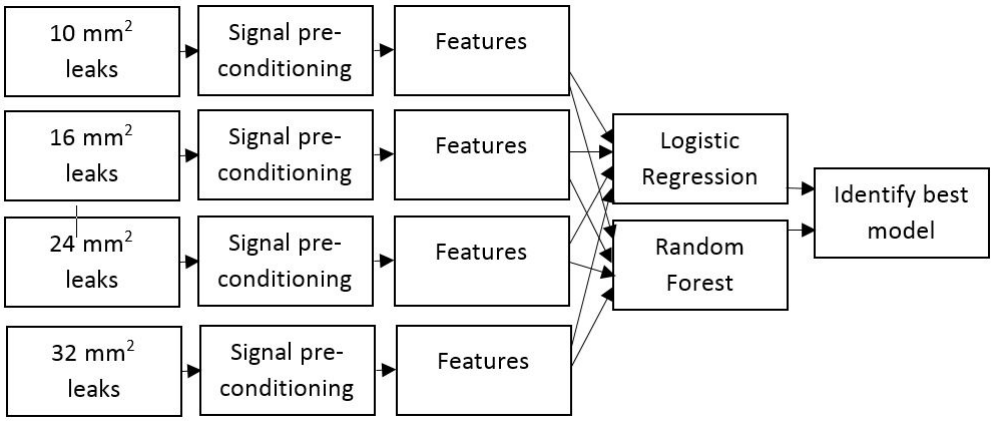


Figure 6.4: *Process flow diagram for leak area prediction.*

6.2.3.2 Linear vs. non-linear models

The use of the correct classifier (linear or non-linear) can heavily influence the accuracy of results. One way to establish if a linear or non-linear classification method should be used is by plotting the sample and assessing the fit to the model. However, as such a large number of features are used in this case, and therefore a significant number of dimensions, it is not possible to plot all feature samples and decide on a model based solely on the visual inspection of those plots. Another method is by training and testing the model with a linear and non-linear classifier and make comparisons by assessing the model accuracy. Table 6.2 presents some initial results whereby a linear classifier (Logistic Regression) and non-linear classifier (Random Forest) are compared. The model was trained initially in a simple form in order to establish the model requirements, by training on all shapes and all 5 leak flow rates but only those under gravel media.

The RF used consists of 1000 decision trees and the entropy splitting criterion was used to measure the quality of a split. The number of features considered when looking for the best decision tree split and the maximum depth of each decision tree was determined by hyperparameter tuning

using 5-fold cross-validation on the training data. Another five-fold cross-validation was used for splitting the training and test data sets, also known as nested cross-validation. During outer five-fold cross-validation all data is split into five equally sized sections and the model is trained on four sections (i.e. 80% of the data) and tested on the remainder section (i.e. 20%). Four additional identical models are each independently trained using a sum of four sections worth of data but each model tested using a different remaining final section. This way all the data was used in training and testing but no individual model is trained on its testing data. The average test accuracy of all five model results in an accuracy score that is almost completely unbiased of how the data was split up into training and testing sets (Varma and Simon, 2006).

Table 6.2 details training and testing on individual shapes (round holes, longitudianl slits and electrofusion joints trained and tested on separate models) and all shapes together (the 'All' data set). The prediction accuracy (%) is used to indicate whether a linear or non-linear classifier is required. The results indicate better model performance when the model is trained and testing on the same leak shape data set, with the 'All' data set nearly always performing poorly in comparison. This suggests that prior knowledge of the leak shape will improve the prediction accuracy of the leak area prediction model. Table 6.2 clearly shows that the non-linear RF model strongly out competes the linear LR model with the RF model having much higher accuracy. A particular distinction is made between the LR and RF models when testing longitudinal slits, whereby the LR and RF models achieve testing accuracy of 55% and 87% respectively. This strongly indicates that the case of prediction of longitudinal slits is a non-linear problem in particular. However, in the case of the round holes, the LR and RF models achieve the same testing accuracy (73%). This is also the case for leak shapes when investigated separately and when investigated as part of the 'All' data set. Therefore, it appears that classification of leak area is generally a non-linear problem.

Table 6.2: *Linear vs non-linear classifiers for leak area prediction.*

Leak shape	Model	Type	Testing accuracy
Electrofusion joint	Logistic Regression	Linear	88%
Electrofusion joint	Random Forest	Non-linear	92%
Round holes	Logistic Regression	Linear	73%
Round holes	Random Forest	Non-linear	73%
Longitudinal Slits	Logistic Regression	Linear	55%
Longitudinal Slits	Random Forest	Non-linear	87%
All Shapes	Logistic Regression	Linear	55%
All shapes	Random Forest	Non-linear	74%

6.2.3.3 Leak area prediction model results

As the investigation above has shown that a non-linear model improved leak area prediction results, an RF model is now trained on all leak area, shapes and backfill types at all 5 leak flow rates and is

shown in Figure 6.5, whereby the results are demonstrated in the form of a confusion matrix. The confusion matrix allows for the visualisation of the model breakdown by accuracy for different leak area sizes. The most accurate predictions of leak area came in the identification of leaks at area measuring 10mm^2 (83% correct prediction), however this had a 17% misclassification rate, whereby 11% of 10mm^2 samples were incorrectly classified as 16mm^2, 3% as 24mm^2 and 3% as 33mm^2. The poorest model predictions were at 24mm^2, which correctly classified only 67% of samples as the correct leak area, but 5% of samples were incorrectly classified as 10mm^2, 14% as 16mm^2, and 14% as 33mm^2. It appears from Figure 6.5 that the smaller leak areas have a higher rate of correct area prediction compared to the larger leak areas. However, overall the RF model correctly identifies leak area to >76% from the VAE signal alone when the whole data set is included (all shapes, all backfill types, all leak areas at all 5 leak flow rates). It is therefore shown here that it is possible to predict leak area from the leak VAE signal to an accuracy of >76%, independent of leak flow rate and backfill type. However, the accuracy rates for different leak area sizes varies depending on the size of the leak, with smaller leaks tending to have higher prediction accuracy.

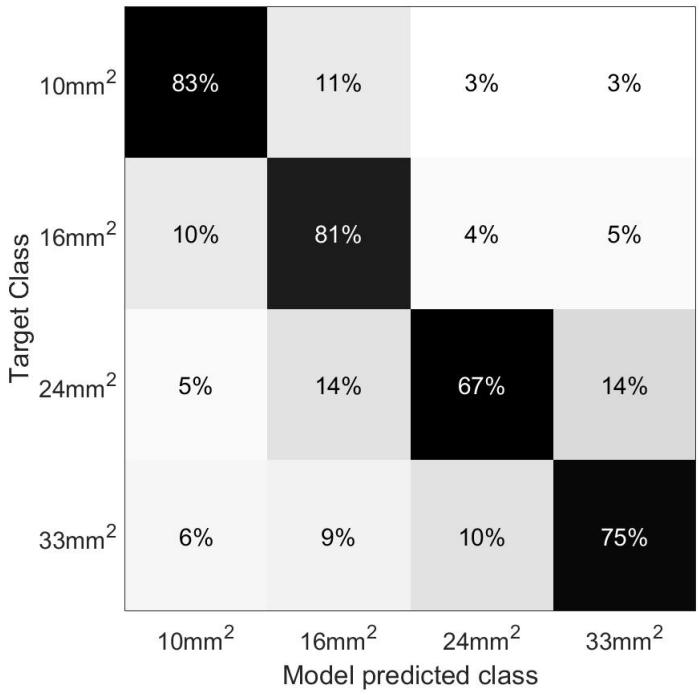


Figure 6.5: *Confusion matrix of leak area prediction rates using the 'All' data set, broken down by leak area prediction accuracy (%).*

The accuracy of the model shown in Figure 6.5 is further broken down by leak flow rate, leak shape and backfill type in Figure 6.6 in order to determine the accuracy of the model at different leak flow rates, leak shapes and backfill types. The accuracy of the model is indicated in terms of it's number of correctly predicted leak areas, shown as a %. The investigation in to the accuracy in predicting individual leak areas at different leak flow rates is shown in Figure 6.6a. Coherent with Figure 6.5, this generally shows better performance at the smaller leak areas, with both 10 mm^2 and 16 mm^2 having the highest leak area prediction accuracy when divided by leak flow rates. There appeared to be a slight trend in that the higher leak flow rates provided more accurate predictions of leak area, which was particularly the case in the larger leak areas (24 mm^2 and 33 mm^2). The most

accurate prediction results however came from the 16 mm^2 at 56 l/min. An investigation in to the prediction accuracy of leak area for the individual leak shapes is shown in Figure 6.6b. Generally, the most accurate leak area predictions in terms of the different leak shapes were from round hole leaks. The round holes are able to provide the most accurate leak area results at diameters of 10 mm^2 and 16 mm^2, with prediction accuracy of approximately 83% and 87% respectively. However, the round holes performed equally the same as the longitudinal slits at 24 mm^2, with a prediction accuracy of approximately 66%. The electrofusion joint was added at 16 mm^2, and was found to have the worst prediction accuracy of 65% correctly estimated leak areas. Interestingly, at the largest leak area of 33 mm^2, electrofusion joints had the highest prediction accuracy in predicting that leak area. In the case of backfill type (Figure 6.6c), it was found that the backfill had a strong influence on the performance of the classifier. In general, it appeared that geotextile fabric and submerged backfill types provided more accurate classifications of leak area at all leak flow rates. However, the most accurate prediction was under a submerged backfill at 56 l/min (approximately 98% prediction accuracy). The lowest prediction accuracy for each backfill type was under gravel media at 44 l/min (approximately 59% accuracy). There did not seem to be a particular observable trend for leak flow rate and backfill, however the middle flow rate (47 l/min) performed poorly for all backfill types apart from Geotextile fabric.

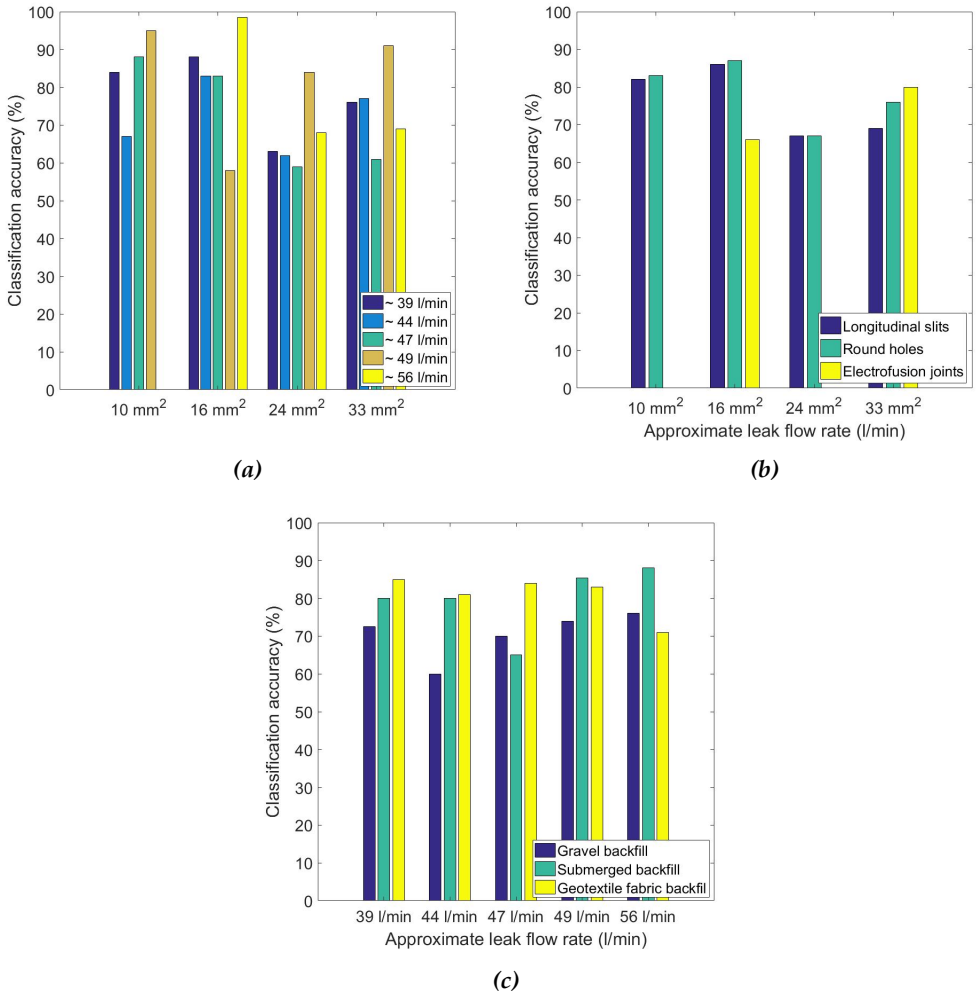


Figure 6.6: *Accuracy in leak area prediction for: (a) leak area for different leak flow rates; (b) the different leak shapes; (c) the different backfill types.*

6.2.3.4 Feature importance

As 22 different features derived from the raw leak signal were used as inputs in to the model, it is of value to understand the most important features to the model as this may provide an explanation as to why some there is variability in prediction accuracy for differing leak area, flow rates, shapes and backfill types. The results of feature importance is shown in Figure 6.7. Figure 6.7 ranks feature importance of the RF model by ranking the use of the 22 different features used as inputs in to the model. This is broken down by the 'All' data set, and also into individual leak shapes. Notably, feature importance in leak area prediction can vary depending on the leak shape and whether or not the 'All' leak shapes data set is considered. The use of the feature RMS of IMF1 appears to be the most important feature spanning all the test cases, but most importantly this was of high value in the 24 mm^2 data set. However, the use of feature importance does not show the effect of different features working in combination with each other and therefore should only be used for indicative purposes.

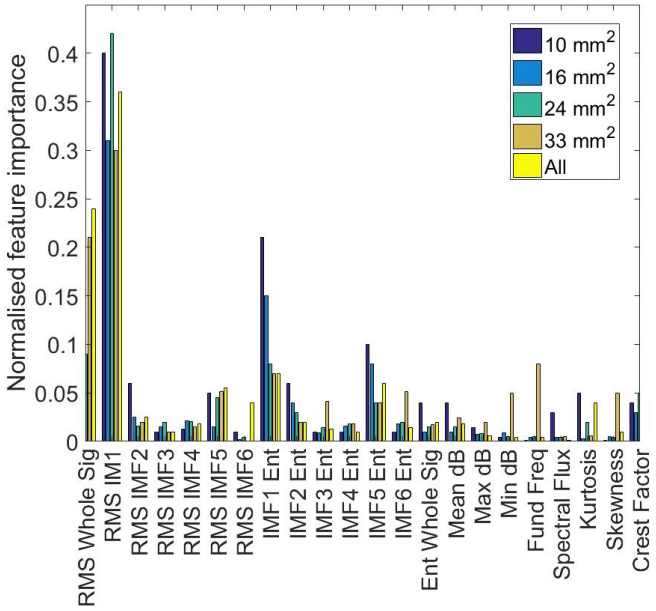


Figure 6.7: *Feature importance in leak area prediction. X-axis represents normalised feature importance.*

6.2.3.5 Sensitivity of the area prediction model to changes in backfill type

It was shown in Chapter 4 that the surrounding backfill can have a strong influence on the leak signal. Figure 6.6c also demonstrated that the backfill influenced the prediction accuracy of the model when considering 'All' data set. Moreover, a large number of backfill types can exist in real WDS, and it is not possible to represent all of these backfills types under laboratory conditions within the remit of this study. However, the sensitivity of the leak area prediction model to changes in backfill can be assessed by training on one backfill type and testing on another. The results for this are shown in Figure 6.8 where the model is trained on gravel backfill and tested on the other two alternative backfill types. This helps to evaluate the importance of backfill type on the models accuracy, and whether or not knowledge *a priori* of the backfill type will improve the models performance. The model clearly shows that there is improved performance when testing on the same backfill type that the model has already been trained on. When testing the model on an alternative backfill that the model was trained on, there is a reduction in classification accuracy.

This appears to be most significant when training the model on gravel backfill and testing on submerged backfill, where there is a reduction in prediction accuracy for all leak shapes. There is a still a reduction in accuracy when testing on geotextile fabric, but this appears to be less so than in submerged backfill. Interestingly, the predictions under electrofusion joints appear to be less sensitive to changes in backfill type than the other two leak shapes. Therefore, it appears that alternate backfills will negatively impact the model (which further depends on leak shape) and therefore data collection and the inclusion of more backfill types within the model training phase will greatly improve the models accuracy and robustness across changing backfill types.

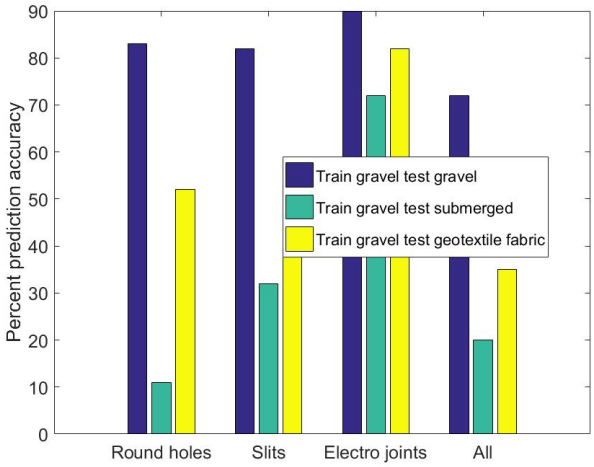


Figure 6.8: *Sensitivity of the leak area prediction model to changes in backfill by training and testing the model on alternative backfill types.*

6.3 (b) Leak shape prediction

6.3.1 Introduction

Chapter 6a has highlighted that the RF model can provide predictions of leak area independent of leak flow rate, leak shape and backfill type. However, of potentially more use to water industry practitioners is the prediction of leak shape from the VAE signal. The aim of this section, following the same format of Chapter 6a, is to derive information leak shape information from the VAE signal. Here, an RF model is trained in order to predict leak shape.

6.3.2 Feature assessment

The signal processing method involved the use of the 22 different features which were extracted from the signal in time and frequency domains. All leak shapes and sizes were decomposed via EEMD generating individual IMF's. These IMF's were transposed in to the frequency domain via the Fourier transform for comparative purposes. The frequency spectrum's of the first six IMF's for a round hole, longitudinal slit and electrofusion joint of equivalent hole area of 32-33.18 mm^2, at leak flow rates of 47-49 l/min are presented in Figure 6.9 .

All IMF's for the different leak shapes studied represent signals of low frequency (<800 Hz) (Figure 6.9). The highest frequency signals within this range are located within the first IMF (IMF1), but this was distinctly low amplitude. Frequency decreases as IMF number increases. The comparison between leak shapes reveals differing frequency distributions across different IMF's depending on leak shape. IMF1 is mainly dominated by signals from the round hole which has the widest spectral band of all the IMF's, the electrofusion joint is largely within IMF's 2, 3 and 4 whilst the longitudinal slit is dominant in IMF5 and 6. The electrofusion joint appears to have greatest power in IMF2, 3 and 4 whilst the Longitudinal slit has more power in IMF5. It is therefore possible that an analysis of the IMF's could provide information on the leak shape. Due to the strong potential of the RMS of IMF1 and IMF2 in predicting leak flow rate (Chapter 5) and leak area (Chapter 6a), the ability of these features to predict leak shape is also assessed. The RMS of IMF1 and IMF2 is shown in Figure 6.10 and Figure 6.11 respectively. It is shown in both of these figures that the leak shape seems to have an influence on the RMS of both IMF1 and IMF2. Both figures appear to indicate that longitudinal slits appear to have a lower RMS of IMF1 when compared to the other leak shapes, whereas round holes and eletrofusion joints have similar values when the leak flow rates are similar. However, whilst there are differences between the leak shapes using this feature, it is incredibly difficult to develop a model to predict leak shape based on this feature alone.

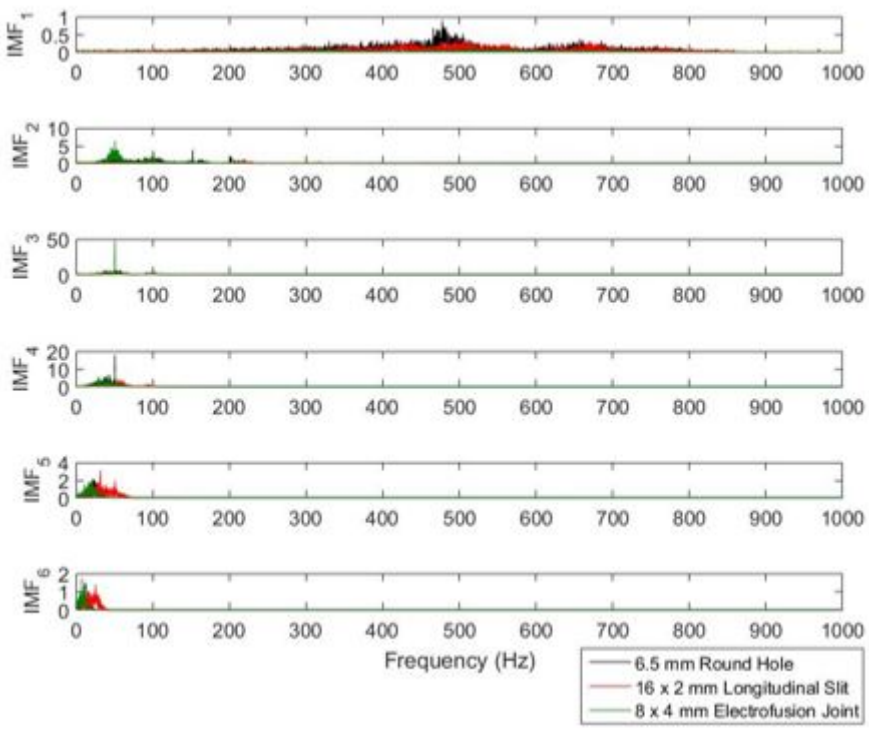


Figure 6.9: *Comparison of leak shapes at individual IMF's following EEMD. All leak flow rates are set between 47-49 l/min and equivalent hole area is 32-33.18 mm^2.*

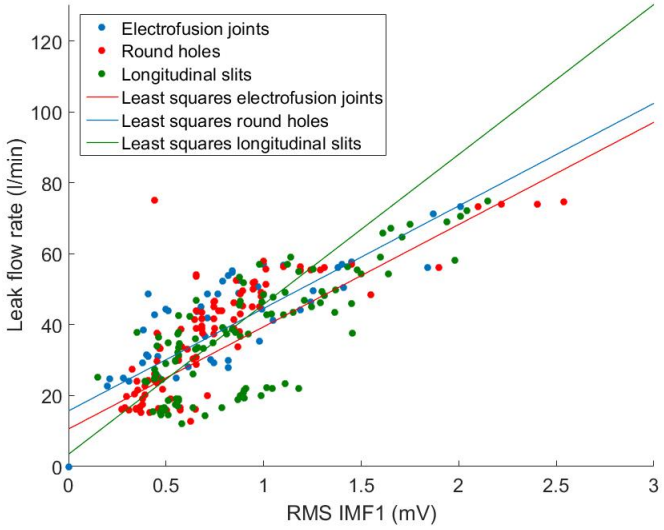


Figure 6.10: *RMS of IMF1 of all leak flow rates with data set divided depending on leak shape, including round holes (R^2= 0.80), longitudinal slits (R^2= 0.96), Electrofusion joints (R^2= 0.76).*

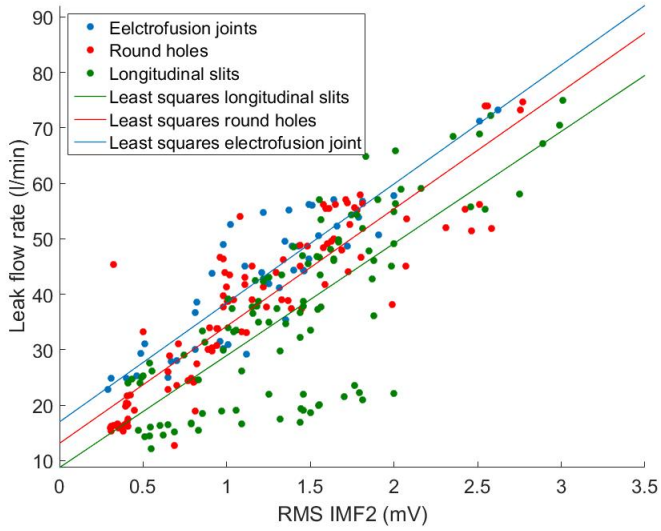


Figure 6.11: *RMS of IMF2 of all leak flow rates with data set divided depending on leak shape, including round holes (R^2= 0.73), longitudinal slits (R^2= 0.55), Electrofusion joints (R^2= 0.91).*

6.3.3 Leak shape prediction model performance

6.3.3.1 Model inputs

Recorded signals were conditioned and pre-filtered as identified in the methods methodology section (Chapter 3), and following this features were derived from the leak signal. A schematic of this process is given in Figure 6.12. Only the 22 features derived from the accelerometer signal were used as inputs to each model (detailed in Table 5.1), i.e. the models were not told the leak shape or the leak flow rate. All leak shapes were initially divided up by their leak area, creating five separate data sets: 10 mm^2; 16 mm^2; 24 mm^2; 33 mm^2; and 'All' (all leak areas used and the model not told the leak area size). As the leak shape corresponds to a matching leak area of another shape, the effect of leak shape is isolated from all other variables. As electrofusion joint data was only available for data set of leak areas 16 mm^2 and 33 mm^2, the data set 10 mm^2 and 24 mm^2 contained less input data.

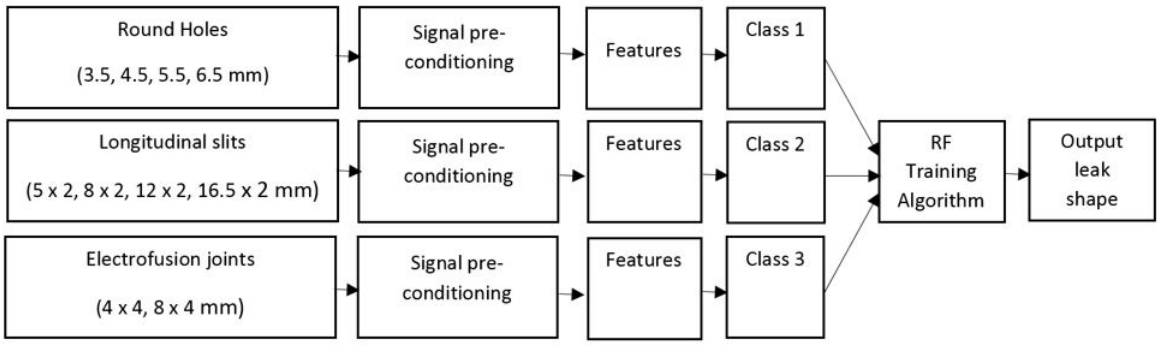


Figure 6.12: *Process flow diagram for leak shape prediction.*

6.3.3.2 Model performance

The performances of the models in classifying leak shape by individual areas is shown in the form of confusion matrices in Figure 6.13. The confusion matrices are further divided by leak area size, the 10 mm^2 data set (Figure 6.13a), 16 mm^2 (Figure 6.13b) data set, 24 mm^2 data set (Figure 6.13c), 33 mm^2 data set (Figure 6.13d) and 'All' leak area sizes (Figure 6.13e). In the case of the 10 mm^2 class, the model correctly classified leak shapes to an average of 87% accuracy. Here, longitudinal sits provided the best classification accuracy at 91%. The introduction of the electrofusion joint shape in the 16 mm^2 class did not have a negative impact on the model, in fact this led to an average prediction accuracy of 89% across all leak shapes (it was previously thought the addition of an extra leak shape would reduce overall classification accuracy, whereas in this case it actually increased classification accuracy by 2%). Interestingly, the elctrofusion joint was the most accurately predicted leak shape within the 16 mm^2 class at 98% of electrofusion joints correctly predicted, with longitudinal slits this time performing the worst in the class (however, this was still a good classification accuracy of 79%). This was also the case when electrofusion joints were included in the 33 mm^2 class, where longitudinal slits had the lowest classification accuracy of 88%. The electrofusion joints had the highest prediction accuracy (93%), and overall classification accuracy for all leak shapes was 90%. In the case of the 24 mm^2 class, the prediction of round holes provided the most accurate estimates, the opposite of the 10 mm^2 class which contained a very similar data set (in terms of the number of leak shapes included). At this hole size, round holes and longitudinal slits were accurately predicted to 99% and 96% respectively. However, in any real system the use of the 'All' dataset is most important as it will not be possible to investigate individual leak area (the leak area would not be known prior to investigations). When classifying with the 'All' data set, the model was able to classify leak shape for all areas at all leak flow rates within all backfill types to $>$81%. When using the 'All' data set, it appeared that classification accuracy was notably higher for round holes compared to the other leak shapes and the longitudinal slit provided the worst predictive performance but this was still accurate at 75%.

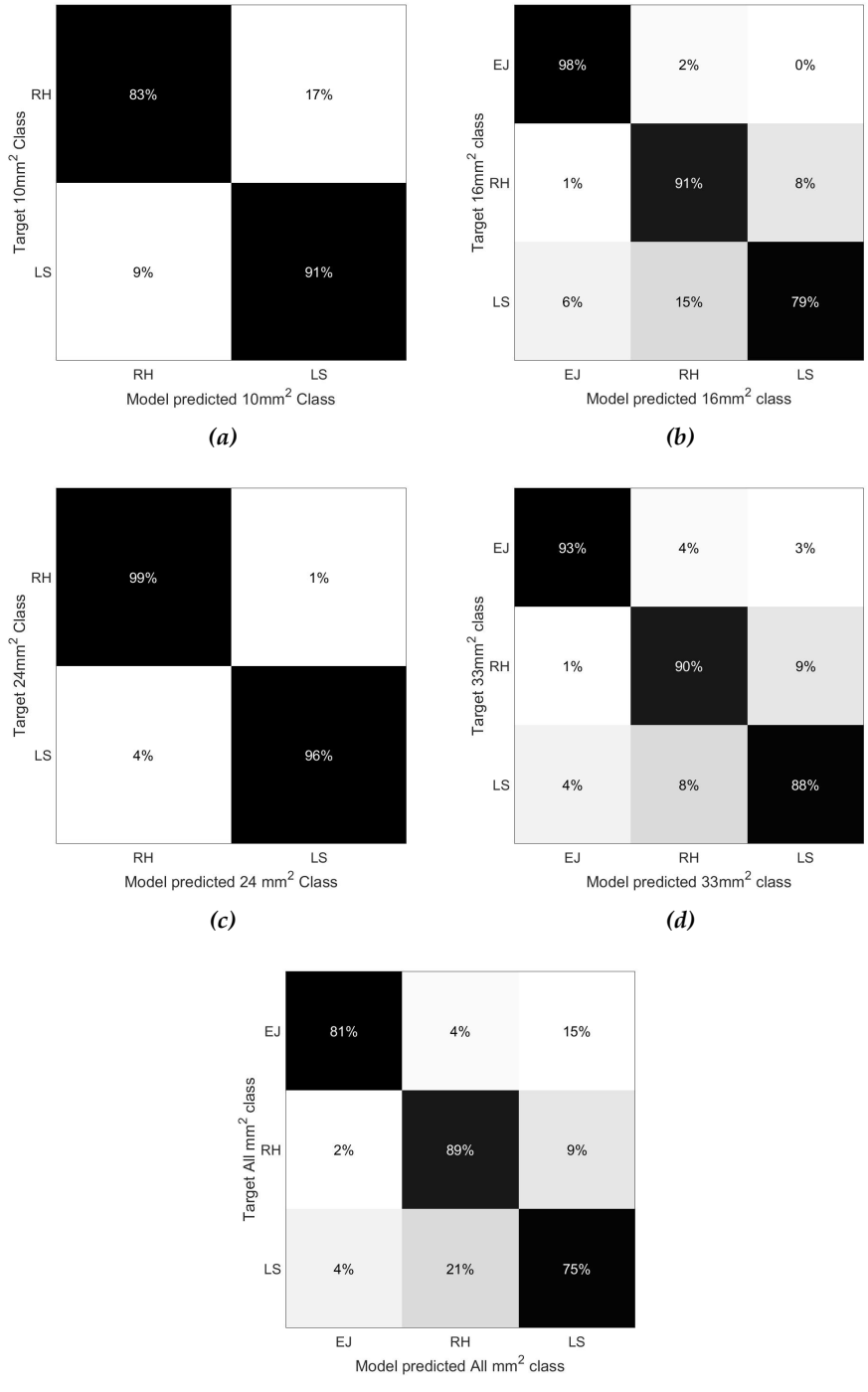


Figure 6.13: *Leak shape prediction model accuracy for each hole shape by leak area: (a) 10 mm^2 class; (b) 16 mm^2 class; (c) 24 mm^2 class; (d) 33 mm^2 class; and (e) 'All' leak areas.*

Figure 6.14 demonstrates the breakdown of the performance of the RF model using the 'All' data set for each leak shape by leak area (Figure 6.14a), leak flow rate (Figure 6.14b) and backfill type (Figure 6.14c). For all leak areas studied, shape classification accuracy is greatest for round holes (>90% accuracy) (Figure 6.14a). The model performs well at 10 mm^2 and 24 mm^2 (>85%), but again this may be due to the fact that these leak areas do not include the electrofusion joint data and therefore there is a smaller number of categories for the model to make predictions. The breakdown of individual leak flow rates for the 'All' data set shows that there is no observable trend

between leak flow rate and shape prediction accuracy (Figure 6.14b). However, the round holes tended to have greater consistency in prediction accuracy at all flow rates. The electrofusion joint performed comparable to the round hole at the lower flow rates, whilst accuracy dropped to <68% at the higher leak flow rates. Generally, prediction of longitudinal slit shape was similar at all leak flow rates, between 68% and 78%. The breakdown of the models performance within individual backfill types shows that at the lowest leak flow rates, classification of leak shape performed better on the gravel media (Figure 6.14c). However, at the highest leak flow rates prediction of leak shape was better under geotextile fabric. The submerged backfill tended to perform worst at the mid to high range leak flow rates.

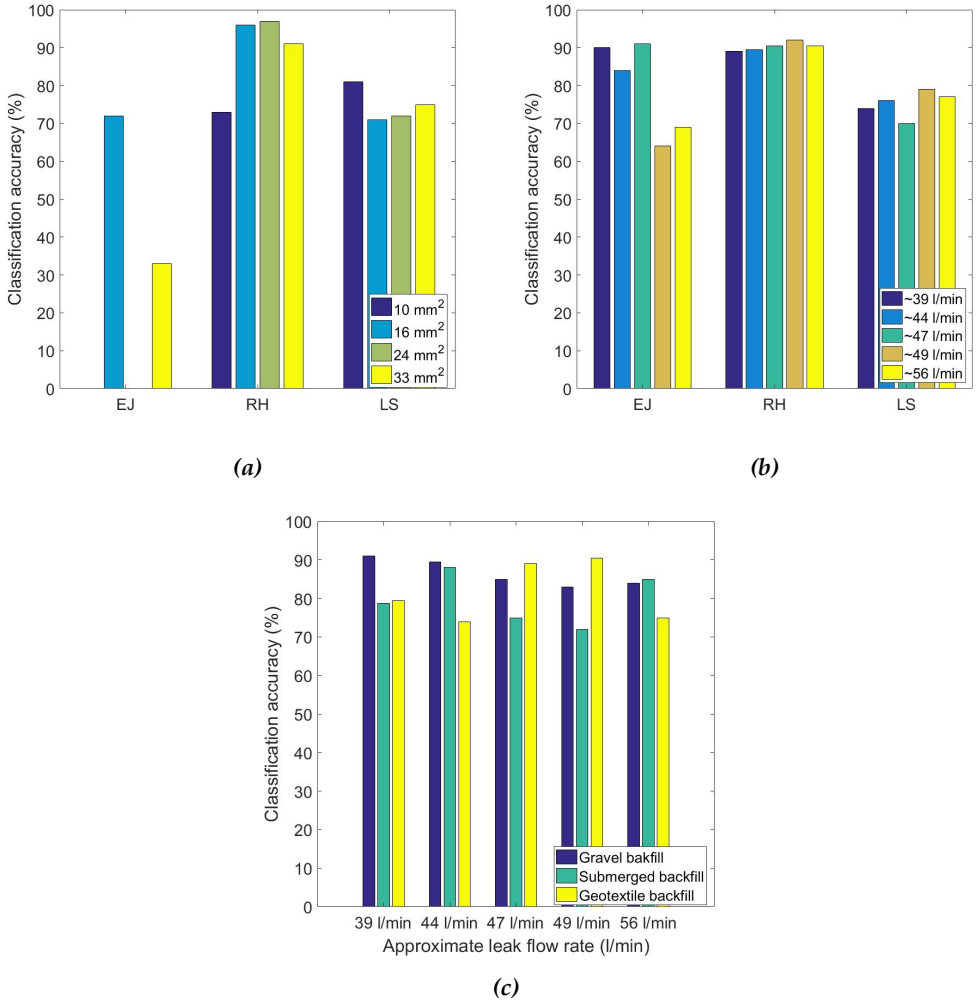


Figure 6.14: *Classification accuracy (%) by different subsets: (a) leak shape by shape area; (b) leak shape by leak flow rate; (c) leak flow rate by media type.*

6.3.3.3 Feature Importance

Figure 6.15 demonstrates feature importance of the RF model by ranking the use of the 22 different features input in to the model. This is broken down by the 'All' data set into individual leak areas and all leak areas. It was found that the most important feature to the leak shape prediction model was the RMS of IMF1. This was true no matter what the leak area and the leak's shape. The model also found other features useful, however the degree to which the model found these features

important depended on whether the model was predicting based on individual leak areas of the 'All' data set.

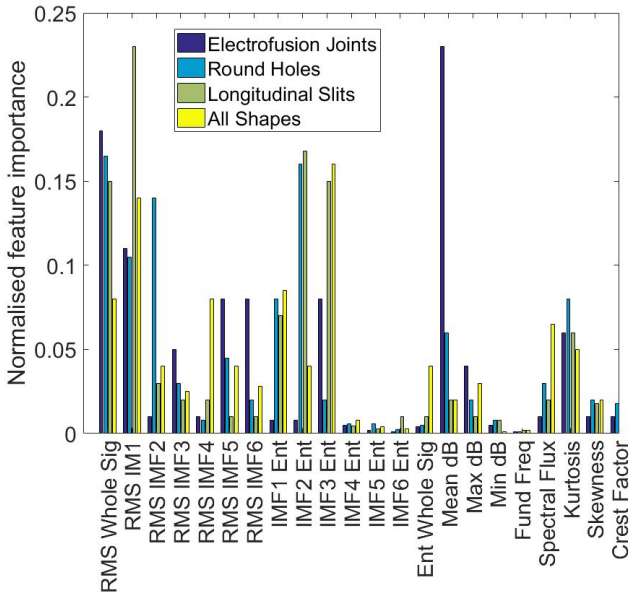


Figure 6.15: *Feature importance when classifying leak shape.*

6.3.3.4 Sensitivity of the leak shape prediction model

In a similar manner to the results presented in Figure 6.8, the effect of backfill type on model performance was evaluated by training and testing on individual backfill types rather than all backfill types. The results for this are shown in Table 6.3. Evidently, backfill type has a significant impact on the performance of leak shape prediction. Overall, training on only one type of backfill and testing the model on a separate backfill type had a largely negative impact on model performance. The worst performance appeared to be training the model on gravel but then testing on submerged data, whereby prediction accuracy was reduced to 21%. Therefore, either backfill type should be known *a priori* or the model needs to be trained and tested on many backfill types to ensure that the model is robust to changes in backfill type.

Table 6.3: *Model accuracy when training and testing on alternate data set.*

Trained on	Tested on	Testing Accuracy
Gravel only	Geotextile	41%
Gravel only	Submerged	21%
Geotextile only	Gravel	36%
Geotextile only	Submerged	32%
Submerged only	Gravel	25%
Submerged only	Geotextile	46%

6.4 Discussion of leak area and leak shape prediction models

6.4.1 Introduction

The research presented herein has demonstrated that the leak's VAE signal contains enough information within it to predict the leak area and shape. A unique experimental investigation used high quality experimental data from various leak shapes of several leak areas at 5 leak flow rates. The aim of this section is to briefly discuss some of the results observed.

6.4.2 Model performance and feature importance

22 features were derived from the leak signal and in combination with a Random Forest model it was possible to predict leak area and leak shape to a relatively high accuracy. The models presented herein provides a robust tool for water industry practitioners due to high classification accuracy at all leak flow rates and all backfill types for both leak area and leak shape prediction. Due to the fact that certain leak shapes have time and pressure dependent growth (Fox, 2016; Ferrante, 2012), knowledge of the leak shape will allow for prioritisation of leak repair. Moreover, the tool provides an opportunity for water companies to collect more data about the shapes of leaks, and thus this information can be linked to further parameters (such as pipe failure mode).

The RMS of IMF1 was found to be the most important feature when classifying leak shape and leak area (Figure 6.15), when using all shapes and leak area within the model. The Fourier spectrum of this feature demonstrates that the amplitude within this IMF is dependent on leak shape (Figure 6.9). Further investigation in to this feature is demonstrated in Figure 6.16, showing the individual RMS of IMF1 for each leak shape and leak area. For all leak shapes it appears generally possible to distinguish between them (of all leak area) individually using this feature. However, in some cases it becomes difficult due to overlap between leak shapes, especially for the electrofusion joint at 32 mm^2. This highlights the necessity for a machine learning based tool to provide the optimum separation between signal features. Leak signal RMS has previously been shown to correlate well with increases in leak flow rate, (Chen et al., 2007; Kaewwaewnoi et al., 2010) and therefore describes information about the leak signal. It was also found by Sun et al. (2016) that the RMS

of each IMF could provide a good descriptor of the leak area. It is therefore logical that the IMF most related to the leak signal and the RMS of this would provide a good method of classifying the leak signal into area and shape categories. However, this feature represents the higher frequency content (350-650 Hz) and when measuring further away from the leak these frequency bands would normally be attenuated due to the pipe acting like a low pass filter (Almeida et al., 2015a). Therefore, this feature will become less effective when moving sensors further away from the leak and therefore less effective in real WDS.

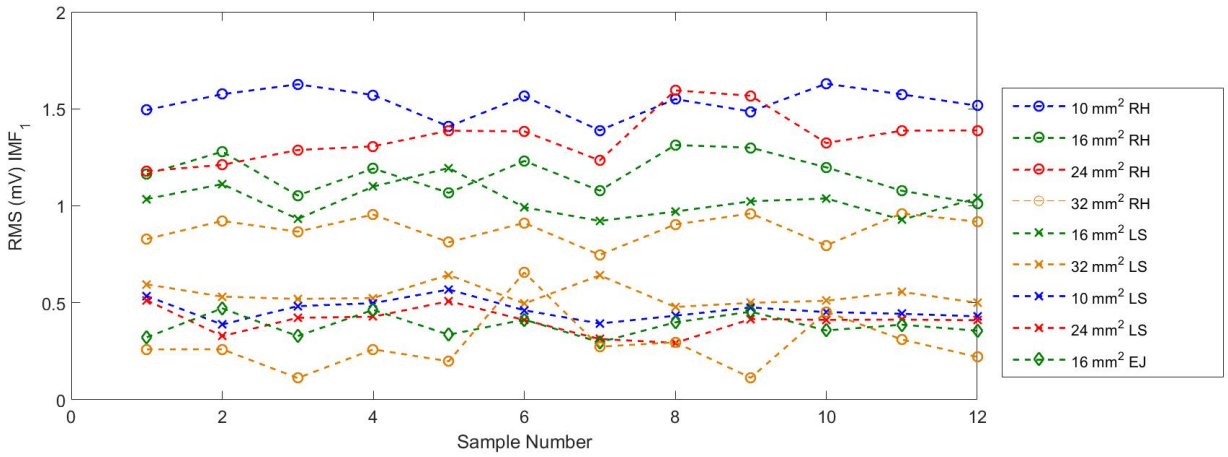


Figure 6.16: *RMS of IMF1 for all leak shapes of different leak area.*

6.4.3 Sensitivity of the models to changes in backfill

The effect of pipe backfill was explored by altering the parameters of both the leak area and leak shape prediction models, training and testing on differing types of backfill (Table 6.3). Backfill type was found to strongly influence the accuracy of the RF models most likely because the leak signal is strongly influenced by the surrounding backfill (Muggleton and Brennan, 2004). Observations of the effect of backfill on the leak signal in Chapter 4 (Figure 4.6) when the leak flow rate is kept consistent between backfill types (47 l/min), showed that the gravel backfill remained higher in amplitude at frequencies >570 Hz compared to the other backfill types, followed by geotextile and then submerged backfill types. Both the gravel and geotextile fabric are similar in that they are a constrained media type (Fox et al., 2016), which will have an impact on the water jet as it leaves the leak hole. However, as it is not possible to achieve good prediction results when training on gravel but testing on geotextile fabric, these results suggest that these media types play a differing role in impacting the generation or transmission of the leak noise. The submerged pipe had the lowest amplitude signals but only at frequencies >570 Hz. A similar effect of backfill was noted by Muggleton and Brennan (2004) who found smaller signal attenuation in the submerged pipe compared to a pipe in soil, however there was a disappearance of the leak signal in the submerged pipe. These results demonstrate the model is more robust when trained on more backfill types, and future work should include a wider range of conditions such as soil saturation, fluidisation and further types which have all been shown to influence leak-media hydraulics (Fox et al., 2016; Van Zyl et al., 2013). Moreover, the impact of these aforementioned variables is still a major research gap in terms of leak VAE signals.

6.4.4 Study limitations

Whilst the tool provided herein provides a tool for water companies to prioritise leak repair, it is not without limitations. A key weakness of this study is the fact that leak shape prediction was undertaken approximately close to the leak. In real-world conditions, it is unlikely that any measurements would take place within such close proximity to the leak. In fact, accelerometers/hydrophones are normally placed on or in nearby fittings, such as valves and hydrants at some distance away. The value of the model has not been tested at further distances, and this remains a key element of future work. However, a number of developments exist in other areas such as pipeline robotics Chatzigeorgiou et al. (2014), which will allow a sensor to travel to a position next to a leak and the system developed here can possibly be integrated into these tools. This study also only addresses a limited number of leak shapes, flow rates, sizes and shapes under only three different backfill types. In real systems, the variety of leaks under varying conditions is huge, and therefore the validity of this system to real world leaks is not known. However, the study has shown that it is possible to differentiate between the leak shapes studied, independent of leak flow rate, leak area and backfill type.

6.5 Chapter summary

This chapter has shown that the leak's VAE signal contains enough information within it to predict leak area and leak shape. 22 features were derived from the leak signal and in combination with a Random Forest model it was possible to predict leak area and shape to a relatively high accuracy. It was also found that the external backfill had a strong impact on the classification accuracy, but training on all backfill types provided a more robust tool with higher classification accuracy in the cases of both models. While the variety of leaks under varying conditions in real-world WDS is huge, and therefore the validity of this system to real world leaks is not known, the proposed technique provided in this paper demonstrates for the first time that it is possible to predict the leak area independent of leak shape, leak flow rate and backfill type. Moreover, it is also possible to predict leak shape independent of leak area, leak flow rate and backfill type. Therefore, this investigative study is the first to demonstrate that there is enough information within a leak signal in order to predict the leak shape. These results are of high importance to the water industry as it provides a tool to predict the leak area and leak shape in order to optimise repair and maintenance strategies.

Chapter 7

Discussion

7.1 Introduction

This thesis has carried out extensive research in to leak VAE signals, including investigations in to the fundamental aspects influencing leak VAE signal characteristics and the development of models to estimate leak flow rate, area and shape. The purpose of this chapter is to discuss the research in the context of its limitations and the future use of the model in real WDS.

7.1.1 Research summary

The research presented in this thesis aimed to derive a method to predict leak flow rate, leak area and leak shape from the VAE signal. This involved preliminary work to develop an understanding of how factors influence the characteristics of leak signals. Several test sections containing artificial leaks representing round holes, longitudinal slits and electrofusion joints of various leak areas were buried within a rectangular open-container and filled with three different backfill types (gravel, geotextile fabric and submerged). All of these parameters were tested at a number of different leak flow rates. Both LS-SVM and RF models are proposed for the prediction of leak flow rate, leak area and leak shape.

This study has managed to accurately predict leak flow rate, leak area, and leak shape with no prior knowledge of other variables (i.e. the prediction is made solely using features derived from the raw signal imported in to the separate models). Despite some limitations associated with the study (which were primarily experimental limitations) discussed in the individual chapters, this research has established the first case in which leak flow rate, leak area and leak shape can be predicted from leak VAE signals in WDS. Moreover, the work presented herein has provided insights in to the leak signal spectra and how differing variables influence the leak signal, shown through the development of a unique experimental methodology. This research explored a major research gap whilst delivering a technique of high value to water companies in order to reduce leakage levels. Individual discussions were given at the end of each chapter. The aim of this section is to discuss the overarching principals of this work and to place the work in to context of the wider industry and the practical application of the demonstrated techniques.

7.1.2 Wider impact of the technique

One of the key advantages of the technique demonstrated in this research is the ease of integration in to existing industry systems. Water companies in the UK use accelerometers and hydrophones regularly in order to detect leaks through leak noise correlation. Therefore, the water companies have the equipment in place already and the integration of the developed technique presented herein within the company systems should be relatively simple. The sensors considered in this study aimed to replicate similar sensors used by the water industry for leak noise correlation, and therefore it should be reasonably easy to use the model as there will be no additional hardware required. However, current leak repair strategies are not always based on leakage volume. In the UK, leak repairs are often prioritised based on customer perception rather than leakage volume. The visible leaks are frequently repaired first due to the importance of customer perception to water companies. However, often this is not the highest volume leakage. The greatest savings in terms of non-revenue water will be a repair prioritisation strategy based on repairing the higher volume leaks first. Therefore, the most difficult task in progressing this research most likely lies with a change in operational management practices and company strategy - which a change may be difficult but most likely to bring the greatest benefit.

A number of authors have attributed much attention to the head-discharge relationship and its association with the pipes condition (May, 1994; Walski et al., 2006), in particular pipe material (Ferrante, 2012; Massari et al., 2012). This relationship is used for both pressure management [for leakage reduction] and transient analysis [for leak detection] (Ferrante et al., 2009). Moreover, the head-discharge relationship is also used with regards to improvements in leakage modelling, as a number of leakage models depend on leak area, as leak area is an extremely important factor influencing the leak behaviour (Clayton and van Zyl, 2007; Ferrante, 2012). Leak area is also an important parameter in the orifice equation and leakage modelling techniques such as the Fixed Area Variable Discharge (FAVAD) model. In the FAVAD model leak area is either fixed or variable, and the leak area is a key parameter. Attempts have also been made in order to produce viscoelastic models in order to predict leak area (Fox et al., 2016), although this needs to be calculate retrospectively. However, the useful and practical tool described in this thesis allows for the rapid characterisation of a leak's shape and area with little effort. Whilst this research is not overly concerned with the factors governing leak hydraulics, the practical tool presented in this thesis can be used to predict leak area for individual leaks and therefore can be used in order to validate existing models or improve leakage models further.

It has been reported that leaks pose a water quality risk (and therefore a risk to human health) as there is potential for contaminant ingress into water distribution systems through leaks (Fox et al., 2014). The level of contamination will involve the size of the leak area (amongst other factors such as driving force) (Fox, 2016). Therefore, a tool which provides the leak area will be useful in judging the risk of contamination due to ingress and therefore the threat to public health. It is therefore anticipated that the results presented in this study will generate strong interest from both academia and industry as water quality is of paramount importance for water companies.

Whilst there are numerous benefits in reducing leakage levels, including environmental, water savings and public health benefits, one of the great benefits and industry motivations for sponsoring this research was for the reduction in water loss in order to reduce revenue loss and meet regulatory requirements in leakage levels. The tool provided herein will help to prioritise leak repair, leading

to cost savings as water companies can direct staff and resources to the largest volume leaks. A significant factor in reporting leakage targets and the position of the water company in regards to SELL, is the correct reporting of the companies leakage levels (Fox, 2016). The DMA water balance approach offers a method which highlights changing trends in DMA scale water patterns (where increasing night-use gives an indication as to the losses within that particular DMA). However, as it provides an indication of total losses, it does not provide an individual leakage rate in cases where there is more than one leak in the network. Any system that manages to predict leak flow rate will be advantageous to water companies, by prioritising leak repair and driving down SELL by repairing the bigger leaks first. This will save more water by fixing less leaks, and costs savings through optimised allocation of company resources.

Minimising leakage levels also strongly relies on a better understanding of the leak's behaviour. For initial reductions, it is relatively easy to reduce leakage levels by simple methods such as pressure management and repairing leaks. However, it becomes more difficult and more expensive to reduce leakage levels further as the remaining leaks are those that are usually harder and more expensive to locate (Ashton and Hope, 2001) (often referred to as background leakage). The SELL mechanism was based on economics, the 'optimum' leakage level is that when the economic expense of leakage control (e.g. via employing leakage operations staff) is equal to the economic expense of the water lost (Ashton and Hope, 2001). This tool presented herein could possibly reduce the SELL value as it will make it cheaper to fix leaks overall, due to the prioritisation of leak repair through the size of the leak. This has numerous benefits to industry and is likely to be an approach favoured by Ofwat due to the overall reduction in leakage levels and follow on impact on reducing customer bills. However, the research presented herein has not addressed cost, precision, practicality of use and the sensitivity to other variables. It is therefore not fully understood how it will impact the SELL value.

Evidently, this tool will allow water companies to better prioritise leak repair based on leak flow rate, area and shape and this can be done simply when a Leakage Technician carries out traditional leak noise correlation. However, water companies use differing strategies for leak repair and this depends on whether or not this is customer driven or leakage priroristation - this tool is for leakage driven and not customer driven. Some priroirstation is based on the risk of the asset.

7.1.3 General study limitations

Whilst the models presented herein provide a tool for water companies to prioritise leak repair, there are a number of limitations associated with this study. A key weakness of this study is the fact that VAE signal recordings was undertaken at close proximity to the leak, at a distance of approximately 30 cm from the leak (at position P_1, Table 3.1). The leak flow rate prediction model presented in Chapter 5 was tested at distances further away from the leak, but only up to a maximum of 8 m from the leak and therefore does not evaluate the model at greater distances from the leak. In real-world conditions, it is unlikely that any measurements would take place within such close proximity to the leak using accelerometers. In fact, accelerometers/hydrophones are normally placed on or in nearby fittings, such as valves and hydrants at some distance away. Whilst the distance between valves and fittings differ, fire hydrants in the UK are normally spaced a minimum of 90 m apart (Regulations, 2017), and therefore the maximum distance the model will be utilised is 90 m. The accuracy of the model has not been tested at further distances, and this remains a key element of future work. However, a number of developments exist in other

areas such as pipeline robotics (Chatzigeorgiou et al., 2014), which will allow a sensor to travel to a position next to a leak - the system developed in herein can possibly be integrated within these tools. Other technologies such as the the Water Research Council Smart Ball (WRC, 2017) and the MTA Pipe Inspector (MTA, 2018), amongst others, 'swim' through the pipe and therefore the sound of a leak can be measured directly next to the leak. Therefore, it is highly possible that the findings generated in this study can be incorporated in to these systems as they also get close to the leak. Moreover, whilst this study has mainly focused on plastic pipe, there was minimal change in signal spectra at distances further away from the leak on the CI and AC pipe materials studied in Chapter 5. Therefore, should the model be trained on CI and AC materials, the model will more likely provide accurate results at greater distances on these materials.

In practice on plastic pipe, the noise will be affected by the transmission distance, reducing in amplitude as it travels further away from the leak. Chapter 4 also alludes to the influence of distance from the leak. Here, there is a loss in amplitude with frequency, and a loss factor per meter (δ/m) is established. The models provided in Chapter 5 and Chapter 6, coupled with this information, is likely to result in more accurate predictions of leak flow rate which reflect the transmission path. However, this is only possible if the effect of transmission distance is known prior to conducting tests. In the case of simple pipe rigs, this is relatively easy to predict based on known equations and loss coefficients (Muggleton and Brennan, 2004). However, in complex WDS the equations describing variations in signal characteristics in response to transmission distance have not yet been shown to be accurate. In Chapter 4, it was shown that the attenuation models were a relatively good predictor at low frequencies and low leak flow rates. Due to the fact that leak signals in plastic pipe tend to be low frequency, the model may well be suitable for plastic pipe. However, the work presented herein has still not provided enough evidence to confirm this as there were contradictory results at higher leak flow rates. It is therefore important that the attenuation characteristics are fully understood before the models provided in this research are fully applicable to real WDS.

This study also only addresses a limited number of leak shapes, flow rates, sizes and shapes under only 3 different backfill types. In real systems, the variety of leaks under varying conditions is huge, and therefore the validity of this system to real world leaks is not known. However, the study has shown that it is possible to differentiate between the leak shapes studied, independent of leak flow rate, leak area and backfill type.

Chapter 8

Conclusions and future work

8.1 Conclusions

The research presented herein aimed to derive the leak flow rate, leak area and leak shape from leak VAE signals in WDS. The research required an initial investigation to develop a fundamental understanding of the factors influencing leak signals, for which a novel experimental methodology took place on a specifically designed state-of-the-art MDPE pipe rig. Unique to this study, for the first time this allowed for the isolation of individual variables that were thought to influence the signal in different ways. This included system pressure, leak flow rate, leak shape, leak area and surrounding backfill type. Some data from a real WDS also allowed for the isolation of pipe material. The experimental investigations showed that of all the factors involved, leak flow rate, leak shape and backfill type had a strong influence governing the characteristics of leak signals when measuring next to the leak. Pipe material only became an important factor when measuring signals at distances away from the leak, and at this position leak signals are a function of leak flow rate, leak shape, backfill and the pipe material due to the transmission of the leak noise. Other factors such as leak area appeared to have less influence on the characteristics of the leak noise.

This study achieved the original aim of developing a tool which could estimate leak flow rate, leak area and leak shape from VAE signals. This was achieved through an exploration of various signal processing and machine learning techniques. This enabled the derivation of a number of features which were found to explain the variation in either leak flow rate, leak area or leak shape. Some of these features are also regularly used in other disciplines, such as speech recognition. Of the several predictive machine learning algorithms explored, LS-SVM models were found to provide the most accurate results to predict leak flow rate. Random Forest models were utilised in order to predict leak shape and leak area, finding accurate performance in predicting leak area and leak shape from VAE signals. The tool developed herein will allow water companies to rapidly prioritise leak repair based on leak flow rate, leak area and leak shape, whilst reducing SELL and improving the efficiency of leakage staff.

8.1.1 Evaluation of research objectives

The conclusions for each research objective (defined originally in Chapter 1) will now be dealt with separately:

1. Review the existing body of literature regarding pipeline leakage detection, characteristics of the leak noise and methods to classify/predict leak flow rate, leak area and leak shape. Also review signal processing and machine learning techniques in order to achieve a classification and prediction model of the aforementioned variables from the leak VAE signal.

   A wide ranging literature review was carried which showed that there is currently no method to predict leak flow rate, leak area and leak shape from leak VAE signals in WDS. The review also highlighted the value of such a tool, in that it could lead to high cost savings for water companies due to reduced non-revenue water due to improved repair prioritisation.

2. Experimentally quantify the factors influencing a leak signal, including: leak flow rate, leak shape, leak area, system pressure, backfill type and pipe material.

   The factors that influence leak signals was evaluated by constructing a unique state-of-the-art pipe rig. This allowed for controlled conditions whereby individual variables could be isolated and their influence on leak signals could be understood. It was found that leak area, leak shape, leak flow rate and backfill type all had a strong influence on the leak signal characteristics. However, leak area appeared to have a negligible impact on the leak signal. A study on the influence of pipe material on leak VAE signals also took place on real, complex WDS and found that pipe material has a significant impact on the leak signal characteristics.

3. Experimentally explore the influence of leaking electrofusion joints on the leak signal, which are more representative of 'real' leaks on plastic WDS.

   The influence of leaky electrofusion joints evaluated by removing part of the joint before electrofusing two pipe sections together. Therefore, a small void remained which allowed for leak to flow out through the voide. It was found that the leaky elctrofusion joint had a significant impact on the characteristics of the leak signal, and in contrast with the round hole and longitudianl slit a very different leak signal was obtained. This means that studies which carry out investigations in to leak signals on plastic pipe using round holes and longitudinal slits are recording leak signals that are not representative of the real leaks found in real WDS.

4. Evaluate and explore different methods of leak signal classification or prediction, in order to classify/predict leak area, leak shape and leak flow rate.

   A number of single feature and multi-feature models were compared in order to evaluate their efficacy in predicting leak flow rate, leak area and leak shape. It was found that using multi-feature models provided improved prediction accuracy and the LS-SVM model provided the overall best model performance (highest accuracy in predictions) in predicting leak flow rate. It was found that leak area and leak shape prediction was best suited to Random Forest models.

5. Design a classification/predictive tool for leak shape and leak flow rate in order to enable water company practitioners to rapidly prioritise leak repair.

A robust tool was developed in order to predict leak flow rate, leak area and leak shape from leak VAE signals. In future work the model should be verified on real leaks to evaluate its efficacy.

8.2 Contributions to knowledge

A number of contributions to knowledge where developed in this thesis, and a brief summary of the key contributions to knowledge are detailed below:

- The research presented herein is the first to establish and carry out experimentally isolate the different factors influencing leak signals. Therefore, the impact of these variables on the leak signal could be established.
- A large and wide ranging data set was collected under controlled conditions when conducting experiments with the pipe rig. This has enabled the collection of one of the largest leak-noise data sets.
- For the first time, a tool to estimate the leak flow rate, leak area and leak shape from leak VAE signals has been developed, which will allow water companies to prioritise leak repair based on these factors. It is possible that this will result in considerable cost savings in terms of the operational expenditure associated with leak repair, and also savings related to reduced non-revenue water (due to more informed repair prioritisation).

8.3 Recommendations for future work

Whilst the presented research designed a robust experimental methodology, the study was not without limitations, which are described and reflected on throughout this thesis. Following the analysis and discussion, the following list proposes future work which will enable to tool to become more robust for use in real WDS. These elements of future work can be divided in to laboratory investigations and investigations on real WDS.

8.3.1 Future work - laboratory investigations

The main recommendations to be undertaken in the laboratory are given below:

- Expansion of experimental data sets to further investigate the influence of sensor distance. Whilst this study evaluated the influence of sensor distance, only limited measurements were taken. In future work the sensors should be moved further away from the leak to test the accuracy of the classification/predictive method to a greater degree than undertaken in this research. The sensor placement should also be evaluated, investigating how placing the sensors on alternate fittings influences the recorded signal and the predictive models.
- Expand the data set to include a higher variation of backfill types under differing conditions to investigate the influence of this on leakage behaviour and VAE signals. Whilst a number of backfill types were investigated in this study, a further programme of work will test more soil types which were unstudied in this work (e.g. clay, various types of sand etc.). The future work should also test differing levels of fluidisation, consolidation and pipe burial depths.

- Further work should expand the amount and variation of leak shapes studied in order to understand the influence of leak shape. Although this study included a number of leak shapes, a more extensive range of leak shapes will allow the shape classification model to become more robust.
- This study only carried out experiments investigating a single leak on each test section used, and therefore the influence of multiple leaks on the model was not evaluated. However, the presence of multiple leaks is likely to alter the received leak signal and therefore this may reduce the accuracy of the leak flow/area/shape prediction model. Therefore the effect of this needs to be evaluated.
- Expand the experimental data set to include a greater variety of valves and fittings not already investigated in this study. It is likely that these factors will influence leak signals, therefore the influence of these factors on model accuracy needs to be evaluated.
- Future laboratory investigations should increase the length of the test rig. In this study, a 26 m pipe rig was evaluated, and sensors moved to various positions throughout the rig. This distance is relatively short compared to the distance measured when leakage staff perform correlations on real WDS. It is recommended that future test rigs are built to represent the maximum spacing distance between fire hydrants (90 m) as this represent the maximum limit for sensor placement in real WDS.
- The influence of pipe material was evaluated in real WDS and a novel experimental methodology isolated the effect of pipe material and leak flow rate. However, future work should alter the pipe material within the pipe rig test section as this will provide more repeatable studies and the comparison between pipe material can be made when other factors (e.g. backfill) are controlled.
- Investigation of real leaks. With careful planning, it may be possible to excavate leaks from real networks and test them in laboratory conditions, by placing them within the test section. This will provide the opportunity to test real leaks in controlled conditions.
- Investigations in to the influence of background noise. Current tests on the pipe rig all took place in controlled conditions, where the level of background noise is minimised. In real WDS, there are a variety of noise sources (such as passing traffic, pump noise, valve noise, rainfall etc.) which can influence leak signals. Future work should aim to replicate the levels of background noise in real systems and assess how this influences leak signals and model prediction accuracy.

8.3.2 Future work - experiments on real WDS

The main recommendations to be undertaken in the laboratory are given below:

- Train and test model on real WDS leaks. Although the study presented herein investigates a number of sized leaky round holes, longitudinal slits and electrofusion joints they do not fully represent real leaks from a real WDS. The model can then be applied to leaks of unknown flow rates and shapes in order to validate the models accuracy.
- Whilst this study moved accelerometers further away from the leak, this was only to a maximum distance of 8 m. Further work should evaluate the influence of distance on the accuracy of

the presented technique. This is especially important on plastic pipes, where the leak signal becomes heavily attenuated when moving sensors further from the leak.

- Carrying out experiments on real WDS is very complex due to the nature of the water network. There are numerous physical discontinuities in the network which can have an influence on the leak signal. This in turn will impact the model accuracy. Whilst this study has outlined the complexity of experiments on real WDS, further work is required in order to establish the influence of these discontinuities on model performance.

- In real WDS, sensors are usually placed at convenient locations such as valves, fittings and fire hydrants. Future work should highlight the importance of the the way in which sensors record a leak signal by evaluating the model efficacy with the sensor placed on various types of fittings that exist in real WDS.

- Implement the leak flow rate, area and hape classification/predictive tool in real WDS and test to evaluate its efficacy on real leaks.

- Determine the impact of the leak flow rate, area and shape classification/prediction tool on the water company reported leakage volumes in order to determine the influence the technique has had on leakage rates. A similar approach can be undertaken to determine the impact on the cost of leak repair.

References

151

Abdulsalam, Hanady, David B Skillicorn, and Patrick Martin. 2007. Streaming Random Forests. In *11th international database engineering and applications symposium (ideas)*. Banff, Alberta, Canada.

Ahadi, Majid, and Mehrdad Sharif Bakhtiar. 2010. Leak detection in water-filled plastic pipes through the application of tuned wavelet transforms to Acoustic Emission signals. *Applied Acoustics* 71(7):634–639.

Almeida, Fabrício, Michael Brennan, Phillip Joseph, Stuart Whitfield, Simon Dray, and Amarildo Paschoalini. 2014a. On the Acoustic Filtering of the Pipe and Sensor in a Buried Plastic Water Pipe and its Effect on Leak Detection: An Experimental Investigation. *Sensors* 14(3):5595–5610.

Almeida, Fabrício C L, Michael J. Brennan, Phillip F. Joseph, Simon Dray, Stuart Whitfield, and Amarildo T. Paschoalini. 2015a. Towards an in-situ measurement of wave velocity in buried plastic water distribution pipes for the purposes of leak location. *Journal of Sound and Vibration* 359:40–55.

Almeida, Fabrício C L, Michael J Brennan1 Phillip F Joseph, Simon Dray, Stuart Witfield, and Amarildo T Paschoalini. 2014b. An investigation into the effects of resonances on the time delay estimate for leak detection in buried plastic water distribution pipes. *Eurodyn 2014: Proceedings of the 9th International Conference on Structural Dynamics* (July):3129–3136.

Almeida, Fabrício César. 2013. Improved Acoustic Methods for Leak Detection in Buried Plastic Water Distribution Pipes. Ph.D. thesis, University of Southampton.

Almeida, Fabrício César, Michael John Brennan, Phillip Frederick Joseph, Simon Dray, Stuart Whitfield, and Amarildo Tabone Paschoalini. 2015b. Measurement of Wave Attenuation in Buried Plastic Water Distribution Pipes. *Strojniški vestnik – Journal of Mechanical Engineering* 60(5):298–306.

Ashton, C H, and V S Hope. 2001. Environmental valuation and the economic level of leakage. *Urban Water* 3:261–270.

AWWA. 1987. Leaks in Water Distribution Systems - A Technical/Economic Overview. Tech. Rep., American Water Works Association., Denver.

Ayala-Cabrera, David, Manuel Herrera, Joaquín Izquierdo, Silvia J. Ocaña-Levario, and Rafael Pérez-García. 2013. GPR-based water leak models in water distribution systems. *Sensors* 13(12): 15912–15936.

BBC. 2018. Thames Water fined £120m over leaks. https://www.bbc.co.uk/news/uk-england-44395763.

Boostan, R., and M.H. Moradi. 2003. Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiological Measurement* 24(2):307–309.

Brabanter, K.De, P Karsmakers, F Ojeda, C Alzate, J De Brabanter, K Pelckmans, B De Moor, J. Vandewalle, and J.A.K. Suykens. 2011. LS-SVMlab Toolbox User's Guide. Tech. Rep., Katholieke Universiteit Leuven.

Breiman, Leo. 2001. Random Forests. *Machine Learning* (45):5–32.

Brennan, M., P. Joseph, J. Muggleton, and Y. Gao. 2006a. Some recent results on the use of acoustic methods to detect water leaks in buried water pipes. *Unpublished research report* (January):1–7.

Brennan, M, P Joseph, J Muggleton, and Y Gao. 2006b. The use of acoustic methods to detect water leaks in buried water pipes. *Unpublished research report* 1–7.

Brown, N. 2007. Intrinsic lifetime of polyethylene pipelines. *Polymer Engineering Science* 47(4): 477–480.

Brunner, Andreas J, and Michel Barbezat. 2007. Acoustic Emission Leak Testing of Pipes for Pressurized Gas Using Active Fiber Composite Elements As Sensors. *Journal of acoustic emission* 25: 42–50.

BSI. 1973. CP 312-1: 1973 Code of practice for plastic pipework (thermoplastics material)- Part 1: General principles and choice of material. Tech. Rep.

Carter, G. Clifford, Albert H. Nuttall, and Peter G. Cable. 1973. The Smoothed Coherence Transform. *Proceedings of the IEEE* 61(10):1497–1498.

Cassa, A., and J.E. van Zyl. 2011. Predicting the head-area slopes and leakage exponents of cracks in pipes. In *Urban water management: Challenges and oppurtunities, ccwi.*, 485–491.

Chatzigeorgiou, Dimitris W., Kamal. Youcef-Toumi, and Rached. Ben-Mansour. 2014. MIT Leak Detector: An in-pipe leak detection robot. In *Ieee international conference on robotics and automation (icra)*. Hong Kong.

Chelabi, M, T Hacib, Y Le Bihan, N Ikhlef, H Boughedda, and M R Mekideche. 2016. Eddy current characterization of small cracks using least square support vector machine. *Journal of Physics D: Applied Physics* 49(15):155303.

Chen, P., P. S K Chua, and G. H. Lim. 2007. A study of hydraulic seal integrity. *Mechanical Systems and Signal Processing* 21(2):1115–1126.

Cist, D.B., and A.E. Schutz. 2001. State of the art for pipe & leak detection: A low-cost GPR Pipe and Leak Detector. Geogmphical Survey Systems Inc. NH 09079(DE-FC26-01NT-I137j. Tech. Rep.

CIWEM. 2017. Water distribution system leakage in the UK: Policy position statement. Tech. Rep.

Clayton, C R I, and J.E. van Zyl. 2007. The effect of pressure on leakage in water distribution systems. *Proceedings of the ICE - Water Management* 160(2):109–114.

Colombo, Andrew F., and Bryan W. Karney. 2002. Energy and Costs of Leaky Pipes: Toward Comprehensive Picture. *Journal of Water Resources Planning and Management* 128(6):441–450.

Demirci, S., E. Yigit, I. H. Eskidemir, and C. Ozedmir. 2012. Ground penetrating radar imaging of water leaks from buried pipes based on back-projection method. *NDT & E International* 35–42.

Duvall, D, and D. Edwards. 2011. Field Failure Mechanisms in HDPE Potable Water Pipe. *ANTEC*.

Elio, J., Q. Crowley, R. Scanlon, J. Hodgson, and S. Long. 2017. Logistic regression model for detecting radon prone areas in Ireland. *Science of the Total Environment* 599-600:1317–1329.

Ferrante, M., B. Brunone, and S. Meniconi. 2009. Leak detection in branched pipe systems coupling wavelet analysis and a Lagrangian model. *Journal of Water Supply: Research and Technology - Aqua* 58(2):95–106.

Ferrante, Marco. 2012. Experimental Investigation of the Effects of Pipe Material on the Leak Head-Discharge Relationship. *Journal of Hydraulic Engineering* 138(8):736–743.

Flandrin, Patrick, Gabriel Rilling, and Paulo Gonçalvés. 2004. Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters* 11(2 PART I):112–114.

Fox, S. 2016. Understanding the Dynamic Leakage Behaviour of Longitudinal Slits in Viscoelastic Pipes. Ph.D. thesis, University of Sheffield.

Fox, S., R. Collins, and J. Boxall. 2016. Physical investigation into the significance of ground conditions on dynamic leakage behaviour. *Journal of Water Supply: Research and Technology - Aqua* 65(2):103–115.

Fox, S., W. Shepherd, R. Collins, and J. Boxall. 2014. Experimental proof of contaminant ingress into a leaking pipe during a transient event. *Procedia Engineering* 70(1):668–677.

Fuchs, H.V., and R. Riehle. 1991. Ten years of experience with leak detection by acoustic signal analysis. *Applied Acoustics* 33:1–19.

Fuller, C.R., and F.J. Fahy. 1982. Characteristics of wave propagation and energy distributions in cylindrical elastic shells filled with fluid. *Journal of Sound and Vibration* 81(4):501–518.

Gao, Y., M. J. Brennan, P. F. Joseph, J. M. Muggleton, and O. Hunaidi. 2005. On the selection of acoustic/vibration sensors for leak detection in plastic water pipes. *Journal of Sound and Vibration* 283(3-5):927–941.

Gao, Y., M.J. Brennan, and P.F. Joseph. 2006. A comparison of time delay estimators for the detection of leak noise signals in plastic water distribution pipes. *Journal of Sound and Vibration* 292(3-5): 552–570.

Gao, Y., M.J. Brennan, P.F. Joseph, and J.M. Muggleton. 2002. Use of Cross-Correlation for Leak Detection in Plastic Pipes. ISVR Technical Memorandum. Tech. Rep. 901.

Gao, Y., M.J. Brennan, P.F. Joseph, J.M. Muggleton, and O. Hunaidi. 2004. A model of the correlation function of leak noise in buried plastic pipes. *Journal of Sound and Vibration* 277(1-2):133–148.

Gao, Yi, and Yanxi Yang. 2013. Crack image detection based on LS-SVM optimized by PSO. In *32nd chinese control conference*, vol. 1, 3602–3606. Xian, China.

Ghazali, M., W. Staszewski, J. Shucksmith, J. Boxall, and S. Beck. 2010. Instantaneous phase and frequency for the detection of leaks and features in a pipeline system. *Structural Health Monitoring* 10(4):351–360.

Ghazali, Mohd Fairusham. 2012. Leak Detection Using Instantaneous Frequency Analysis. Ph.D. thesis, University of Sheffield.

Gneiting, Tilmann. 2011. Making and Evaluating Point Forecasts. *Journal of the American Statistical Association* 106(494):746–762. 0912.0902.

Greyvenstein, Bruce, and J.E.V. Zyle. 2004. An Experimental Investigation Into the Pressure-Leakage Relationship of Some Failed. In *Water loss proceedings*, 2004. November.

Grosse, C.U., and M. Ohtsud. 2008. *Acoustic Emission Testing*. 1st ed. 1, Berlin, Germany.: Springer-Verlag Berlin Heidelberg.

Guardian. 2017. Water companies losing vast amounts through leakage, as drought fears rise. https://www.theguardian.com/environment/2017/may/11/water-companies-losing-vast-amounts-through-leakage-raising-drought-fears.

Guyon, Isabelle. 2003. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* 3:1157–1182.

Hassan, Hebatallah, Amr Badr, and M B Abdelhalim. 2015. Prediction of O-glycosylation Sites Using Random Forest and GA-Tuned PSO Technique. *Bioinformatics and Biology Insights* 9:103–109.

Hauke, Jan, and Tomasz Kossowski. 2011. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. *Quaestiones Geographicae* 30(2):87–93.

Huang, N E, Z Shen, S R Long, M C Wu, H H Shih, Q Zheng, N.-C. Yen, C C Tung, and H H Liu. 1998. The empirical mode decomposition and the {H}ilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Royal Soc. Lond.* A 495:903–995.

Huang, Norden E., and S.S. Shen. 2005. *Hilbert-Huan Transform and Its Applications*. Singapore: World Scientific.

Humphrey, N., M. Loveday, and S. Toombs. 2012. Leak Detection on Plastic Pipes 1–101.

Hunaidi, Osama, and Wing T Chu. 1999. Acoustical characteristics of leak signals in plastic water distribution pipes. *Applied Acoustics* 58:235–254.

Hunaidi, Osama, Alex Wang, Marc Bracken, Tony Gambino, and Charlie Fricke. 2004. Acoustic Methods for Locating Leaks in Municipal Water Pipe Networks. 1–14. Dead Sea - Jordan: International Water Demand Management Conference.

Ionel, R., S. Ionel, and A Ignea. 2010. Improved leak detection quality by automatic signal filtering. In *Ieee international joint conferences on computational cybernetics and technical informatics (iccc-conti)*. Timisora, Romania.

Kaewwaewnoi, W., A. Prateepasen, and P. Kaewtrakulpong. 2010. Investigation of the Relationship Between Internal Fluid Leakage through a Valve and the Acoustic Emission Generated from the Leakage. *Measurement: Journal of the International Measurement Confederation* 43(2):274–282.

Kaewwaewnoi, Watit, Asa Prateepasen, and Pakorn Kaewtrakulpong. 2007. A Study on Correlation of AE Signals from Different AE Sensors in Valve Leakage Rate Detection. In *Ecti transactions on electrical eng. electronics, and communications.*, 113–117. August.

Khawaja, Taimoor, and George J. Vachtsevanos. 2009. A novel Bayesian least squares support vector machine based anomaly detector for fault diagnosis. *Proceedings of the Annual Conference of the Prognostics and Health Management Society* 1–8.

Khodayari-Rostamabad, A., J.P. Reilly, N.K. Nikolova, J.R. Hare, and S. Pasha. 2009. Machine Learning Techniques for the Analysis of Magnetic Flux Leakage Images in Pipeline Inspection. *IEEE Transactions on Magnetics* 45(8):1–14.

Khulief, Y. a., A. Khalifa, R. Ben Mansour, and M. a. Habib. 2012. Acoustic Detection of Leaks in Water Pipelines Using Measurements inside Pipe. *Journal of Pipeline Systems Engineering and Practice* 3(2):47–54.

Knapp, C. H., and G. C. Carter. 1976. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing* 24(4):320– 327.

Landau, H. J. 1967. Sampling, Data Transmission, and the Nyquist Rate. *Proceedings of the IEEE* 55(10):1701–1706.

Landis. 2016. Soilscapes. http://www.landis.org.uk/soilscapes/.

Langley, P. 1994. Selection of relevant features in machine learning. In *Proceedings of the aaai fall symposium on relevance*. AAAI Press.

Lechevallier, Mark W, Richard W Gullick, Mohammad R Karim, Melinda Friedman, and James E Funk. 2003. The potential for health risks from intrusion of contmainants into the distribution system from pressure transients. *Journal of Water and Health* 3–14.

Lewis, P., H. Fitriani, and I. Arocho. 2015. Engine variable impact analysis of fuel use and emission for heavy-duty diesel maintenance equipment. *Transportation Research Record: Journal of the Transportation Research Board* 2482(02):8–15.

Long, R. 2001. Monitoring acoustic wave propagation in buried cast iron water pipes. *AIP Conference Proceedings* 557(2002):1202–1209.

Louppe, Gilles. 2014. Understanding Random Forests: From Theory to Practice (July). `1407.7502`.

Mandic, Danilo P., Naveed Ur Rehman, Zhaohua Wu, and Norden E. Huang. 2013. Empirical mode decomposition-based time-frequency analysis of multivariate signals: The power of adaptive data analysis. *IEEE Signal Processing Magazine* 30(6):74–86.

Marple, Lawrence. 1987. *Digital spectral analysis: with applications*. London: Prentice-Hall International.

Mashford, John, Dhammika De Silva, Donavan Marney, and Stewart Burn. 2009. An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine. *NSS 2009 - Network and System Security* 534–539.

Massari, C., M. Ferrante, B. Brunone, and S Meniconi. 2012. Is the leak head–discharge relationship in polyethylene pipes a bijective function? *Journal of Hydrualic Research IAHR* 50(4):409–417.

May, J. 1994. Pressure dependent leakage. World Water and Environmental Engineering.

Mba, D. 2003. Acoustic Emissions and monitoring bearing health. *Tribiology Transactions* 46(3): 447–451.

McMahon, W., M.H. Burtwell, and M. Evans. 2005. Minimising street works disruption: the real costs of street works to the utility industry and society, Tech. Rep. 05/WM/12/8. Tech. Rep., Water Industry Research, London, UK.

MedCalc. 2017. Logistic regression.

Meland, E., V. Henriksen, E. Hennie, and M. Rasmussen. 2011a. Spectral analysis of internally leaking shut-down valves. *Measurement: Journal of the International Measurement Confederation* 44(6): 1059–1072.

Meland, E., N.F. Thornhill, E. Lunde, and M. Rasmussen. 2011b. Quantification of valve leakage rates. *AIChE Journal* 58(4):1181–1193.

Meyers, G., Z. Kepelan, and E. Keedwell. 2017. Short-term forecasting of turbidity in trunk main networks. *Water Research* 67–79.

Microsoft. 2017. How to choose algorithms for Microsoft Azure Machine Learning.

Miller, A.J. 1990. *Subset selection in regression*. Chapman and Hall.

Miller, R.K, a.a Pollock, D.J Watts, J.M Carlyle, a.N Tafuri, and J.J Yezzi. 1999. A reference standard for the development of acoustic emission pipeline leak detection techniques. *NDT & E International* 32(1):1–8.

Mohamad, Edy Tonnizam, Danial Jahed Armaghani, Amir Mahdyar, Ibrahim Komoo, Khairul Anuar Kassim, Arham Abdullah, and Muhd Zaimi Abd Majid. 2017. Utilizing regression models to find functions for determining ripping production based on laboratory tests. *Measurement: Journal of the International Measurement Confederation* 111(January):216–225.

Mostafapour, A., and S. Davoudi. 2013. Analysis of leakage in high pressure pipe using acoustic emission method. *Applied Acoustics* 74(3):335–342.

MTA. 2018. MTA Pipe Inspector. http://www.mta-messtechnik.at/mzga/pipe-inspector/?lang=en.

Muggleton, J. M., and J. Yan. 2013. Wavenumber prediction and measurement of axisymmetric waves in buried fluid-filled pipes: Inclusion of shear coupling at a lubricated pipe/soil interface. *Journal of Sound and Vibration* 332(5):1216–1230.

Muggleton, J.M., and M.J. Brennan. 2004. Leak noise propagation and attenuation in submerged plastic water pipes. *Journal of Sound and Vibration* 278(3):527–537.

Muggleton, J.M, and M.J Brennan. 2008. The design and Instrumentation of an experimental rig to investigate acoustic methods for the detection and location of underground piping systems. *Applied Acoustics* 69:1101–1107.

Muggleton, J.M., M.J. Brennan, and Y. Gao. 2006. A novel sensor for measuring the acoustic pressure in buried plastic water pipes. *Journal of Sound and Vibration* 295:1085–1098.

Muggleton, J.M., M.J. Brennan, and P.W. Linford. 2004. Axisymmetric wave propagation in fluid-filled pipes: wavenumber measurements in in vacuo and buried pipes. *Journal of Sound and Vibration* 270(1-2):171–190.

Muggleton, J.M., M.J. Brennan, and R.J. Pinnington. 2002. Wavenumber Prediction of Waves in Buried Pipes for Water Leak Detection. *Journal of Sound and Vibration* 249(5):939–954.

Murvay, Pal Stefan, and Ioan Silea. 2012. A survey on gas leak detection and localization techniques. *Journal of Loss Prevention in the Process Industries* 25(6):966–973.

Nishimura, H., A. Nakashiba, M. Nakakura, and K. Sasai. 1993. Fatigue behavior of medium-density polyethylene pipes for gas distribution. *Polymer Engineering and Science* 13(14):895–900.

Ofwat. 2010. Service and delivery performance of the water companies in England and Wales 2009-10 report. http://www.ofwat.gov.uk/regulating/reporting/rpt los 2009-10supinfo.pdf.

Pal, Maninder. 2008. Leak detection and location in polyethylene pipes. Ph.D. thesis.

Pal, Maninder, Neil Dixon, and James Flint. 2010. Detecting & Locating Leaks in Water Distribution Polyethylene Pipes. *World Congress on Engineering* II.

Palczewska, Anna, Jan Palczewski, Richard Marchese Robinson, and Daniel Neagu. 2013. Interpreting random forest models using a feature contribution method. In *Ieee 14th international conference on information reuse and integration*, 112–119. San Francisco, CA, USA.

Papastefanou, Anastasia. 2011. An experimental investigation of leak noise from water filled plastic pipes. Ph.D. thesis.

Primayer. 2016. Leak Sizer. *Business* 44(0):6–7.

Prime, MB, and D.W. Shevitz. 1996. Linear and nonlinear methods for detecting cracks in beams. *Proceedings pages* 836.

Puust, R., Z. Kapelan, D. A. Savic, and T. Koppel. 2010. A review of methods for leakage management in pipe networks. *Urban Water Journal* 7(1):25–45.

Qing, Mao, Zhang Jinghui, Luo Yushan, Wang Haijun, and Duan Quan. 2006. Experimental studies of orifice-induced wall pressure fluctuations and pipe vibration. *International Journal of Pressure Vessels and Piping* 83(7):505–511.

Reed, Dw, Gf Read, and M Price. 1989. Discussion: the Influence of Mains Leakage and Urban Drainage on Ground Water Levels Beneath Conurbations in the Uk. *ICE Proceedings* 86(5):1019–1020.

Reethof, G. 1978. Turbulence-Generated Noise in Pipe Flow. *Annual Review of Fluid Mechanics* 10(1):333–367.

Regulations, The Building. 2017. Section 15 Fire mains and hydrants.

Rodgers, J.L., and W Alan Nicewander. 1988. Thirteen Ways to Look at the Correlation Coefficient. *The American Statistician* 42(1):59–66.

Ross, D. 1976. *Mechanisms of underwater noise*. Ross, D. Mechanics of underwater noise. Pergamon Press.

RPS. 2008. Providing Best Practice Guidance on the Inclusion of Externalities in the ELL Calculation. Tech. Rep.

Ruchonnet, Nicolas, Sébastien Alligné, Christophe Nicolet, and François Avellan. 2012. Cavitation influence on hydroacoustic resonance in pipe. *Journal of Fluids and Structures* 28:180–193.

Sadeghioon, Ali, Nicole Metje, David Chapman, and Carl Anthony. 2014. SmartPipes: Smart Wireless Sensor Networks for Leak Detection in Water Pipelines. *Journal of Sensor and Actuator Networks* 3(1):64–78.

Schmid, H. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *The international conference on new methods in language process*. Manchester, UK.

Severn Trent, Water. 2017. We're detecting leaks from space! https://www.stwater.co.uk/news/news-releases/wearedetectingleaksfromspace/.

Shardlow, Matthew. 2002. An Analysis of Feature Selection Techniques. Ph.D. thesis.

Shehadeh, M, J a Steel, and R L Reuben. 2006. Acoustic Emission Source Location for Steel Pipe and Pipeline Applications: The Role of Arrival Time Estimation. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering* 220(2):121–133.

Shen, Jichen, Hongfei Chang, and Yang Li. 2011. Pressure vessel state investigation based upon the least squares support vector machine. *Mathematical and Computer Modelling* 54(3-4):883–887.

Sheng, J., S. Dong, Z. Liu, and H. Gao. 2016. Fault feature extraction method based on local mean decomposition Shannon entropy and improved kernel principal component analysis model. *Advances in Mechanical Engineering* 8(8):1–8.

Shucksmith, J., J.B. Boxall, W.J. Staszewski, A. Seth, and S.B.M. Beck. 2012. Onsite leak location in a pipe network by cepstrum analysis of pressure transients. *Journal of the American Water Works Association* 104(8):35–36.

Si, Lei, Zhongbin Wang, Xinhua Liu, Chao Tan, Ze Liu, and Jing Xu. 2016. Identification of Shearer Cutting Patterns Using Vibration Signals Based on a Least Squares Support Vector Machine with an Improved Fruit Fly Optimization Algorithm. *Sensors* 16(1):90.

Singh, Manjeet, and Manjeet Singh. 2014. Audio Noise Reduction Using Wavelet Types With. *International journal of computer and organisational trends* 6(1):86–89.

Statsoft. 2017. Partial Least Squares Regression. http://www.statsoft.com/Textbook/Partial-Least-Squares.

Suez. 2015. Advanced leakage detection using Helium gas. http://www.swig.org.uk/wp-content/uploads/2015/01/Steve-George-2015.pdf.

Sun, Jiedi, Qiyang Xiao, Jiangtao Wen, and Ying Zhang. 2016. Natural gas pipeline leak aperture identification and location based on local mean decomposition analysis. *Measurement: Journal of the International Measurement Confederation* 79:147–157.

Suykens, J. A.K, and J. Vandewalle. 1999. Least squares support vector machine classifiers.

Taub, H., and D.L. Schilling. 2002. *Principles of Communication Systems*. 2nd ed. New York: The Mcgraw Hill Inc.

Tayefi, Pedrom. 2014. The fatigue response of electrofusion joins when subject to contamination. Ph.D. thesis, University of Sheffield.

ThamesWater. 2018. Understanding water pressure. https://www.thameswater.co.uk/Help-and-Advice/No-Water-or-low-pressure/Unde.

Thompson, M., D.J. Allwright, C.J. Chapman, S.D. Howison, and J.R. Ockendon. 2001. Noise generation by water pipe leaks. Study report of the 40th European Study group with industry. Tech. Rep.

UKWIR. 2008. National Sewers and Water Mains Failure Database (08/RG/05/26). Tech. Rep.

UtilityWeek. 2018. Thames Water returns £120m to customers for leakage failure.

Van Zyl, J E, M O a Alsaydalani, C R I Clayton, T Bird, and A Dennis. 2013. Soil fluidisation outside leaks in water distribution pipes – preliminary observations. *Water Management* 166(WM10): 546–555.

Varma, Sudhir, and Richard Simon. 2006. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7(1):91.

Ver, Istvan, Beranek, Leo. 2005. Noise and Vibration Control Engineering: Principles and Applications, 2nd Edition. In *Noise and vibration control engineering: Principles and applications, 2nd edition*, 976.

Walski, T., W. Bezts, E. Posluszny, and M. Weir. 2006. Modelling leakage reduction through pressure control. *AWWA* 98(4):147–155.

Wassaf, W.A., M.N. Bassim, M. Houssny-Emam, and K. Tangri. 1985. Acoustic emission spectra due to leaks from circular holes and rectangular slits. *The Journal of the Acoustical Society of America* 77(3):916–923.

Water, Briefing. 2017. Anglian Water trials thermal-imaging drones for leak detection.

Water Service Association of Austrialia. 2003. Common failure modes in pressurised pipeline systems.

Water Services Association of Australia. 2012. Failure Modes in Pressurised Pipeline Systems. Tech. Rep. August.

WRC. 2017. WRc carries out successful SmartBall and PipeDiver Surveys for Caledonia Water Alliance on behalf of Scottish Water. http://www.wrcplc.co.uk/wrc-carries-out-successful-smartball-and-pipediver-surveys-for-caledonia-water-alliance-on-behalf-of-scottish-water.

Wu, Zhaohua, and Norden E. Huang. 2005. Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method. *Advances in Adaptive Data Analysis* 01(01):1.

WWI. 2017. UK water utilities under fire for unimprove leakage levels.

Yan, Jin, Yang Heng-hu, Yang Hong, Zhang Feng, Liu Zhen, Wang Ping, and Yang Yan. 2015. Nondestructive Detection of Valves Using Acoustic Emission Technique 2015.

Zhan, Xiaobin, Shulan Jiang, Yili Yang, Jian Liang, Tielin Shi, and Xiwen Li. 2015. Inline Measurement of Particle Concentrations in Multicomponent Suspensions using Ultrasonic Sensor and Least Squares Support Vector Machines. *Sensors* 15(9):24109–24124.

Zhang, Cha, Dinei Florêlncio, and Zhengyou Zhang. 2008. WHY does PHAT work well in low noise, reverberative environments? *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (March):2565–2568.

Zhang, Jian, Ruqiang Yan, Robert X. Gao, and Zhihua Feng. 2010. Performance enhancement of ensemble empirical mode decomposition. *Mechanical Systems and Signal Processing* 24(7):2104–2123.

Zhao, W.F. 2017. Automatic recognition of loess landforms using Random Forest method. *Journal of Mountain Science* 14(5):885–897.

Zhou, Haomiao, Zhihong Deng, Yuanqing Xia, and Mengyin Fu. 2016. A new sampling method in particle filter based on Pearson correlation coefficient. *Neurocomputing* 216:208–215.

Zorman, Milan, Milojka Molan Stiglic, Peter Kokol, and Ivan Maltic. 1997. The Limitations of Decision Trees and Automatic Learning in Real World Medical Decision Making 21(6):403–415.