DOCTOR OF PHILOSOPHY

PIPELINE LEAK AND CRACK IDENTIFICATION USING ACOUSTIC EMISSION TECHNIQUES
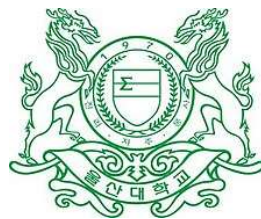
**The Graduate School
of the University of Ulsan
Department of Electrical, Electronic and
Computer Engineering**

BUI QUY THANG

Pipeline Leak and Crack Identification using Acoustic Emission Techniques

Supervisor: Prof. Kim, Jong–Myon

A Dissertation

Submitted to
the Graduate School of the University of Ulsan
In partial Fulfillment of the Requirements
for the Degree of
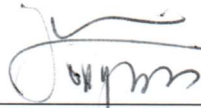
Doctor of Philosophy

by

Bui Quy Thang

Department of Electrical, Electronic
and Computer Engineering,
University of Ulsan, Korea
February 2022

Pipeline Leak and Crack Identification using Acoustic Emission Techniques

This certifies that the dissertation
of Bui Quy Thang is approved.

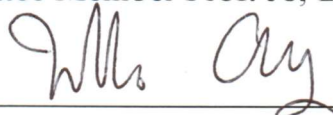Committee Chair Prof. Kwon, Young–Keun

---

Advisor Prof. Kim, Jong–Myon

Committee Member Prof. Jo, Dong–Sik

Committee Member Prof. Chung, Jin–Ho

Committee Member Prof. Kim, Cheol–Hong

Department of Electrical, Electronic
and Computer Engineering,
University of Ulsan, Korea
February 2022

VITA

Bui Quy Thang was born in Hai Duong, Vietnam, in 1985. He completed his B.Sc. in Electrical and Electronic Engineering in 2009, and M.Sc. in Cybernetics and Automation Engineering in 2013 from Le Quy Don Technical University, Hanoi, Vietnam. He has been working as a lecturer in the Department of Electronic Technology, Institute of System Integration, Le Quy Don Technical University, Vietnam since he received the B.Sc. Degree. Since March 2018, he has been pursuing his Ph.D. Degree in the Department of Electrical, Electronic and Computer Engineering, University of Ulsan, South Korea under the supervision of Professor Jong–Myon Kim. His Ph.D. dissertation is based on exploring advanced techniques of signal processing and artificial intelligence for fault diagnostics of pipelines using acoustic emission. His current research interests include algorithms of signal detection, pattern recognition, system modelling, optimal estimation, control and automation, radio technologies, and embedded systems.

DEDICATION

This thesis is dedicated to

many beautiful things that

inspired us

From who we were

to who we are

and whom we will become

We are traveling

on a journey of discovery

full of curiosity

To create

many beautiful things that

inspire others

This effort is dedicated to the two people who have taught me the most important lessons in my life. To my wife, whose love is the greatest gift of my life and whose commitment, support and patience are true models for all; and to my son, who inspires me to grow and who will always be a continual source of pride and enlightenment.

Thanks, mom and dad for always believing in me and for encouraging me to strive for my dreams.

ACKNOWLEGEMENTS

Bui Quy Thang

University of Ulsan, Ulsan, South Korea

February 2022

ABSTRACT

The health monitoring of pipelines has become extremely essential since pipeline networks were widely constructed. The primary aim of this process is to pinpoint any abnormality early in a pipeline to require maintenance or replacement as soon as possible because a defective pipe that has not been repaired would cause serious consequences for nearby nature and human. Pipeline failures can mainly be classified into a range of types including leaks, cracks, and corrosions. Cracks and corrosions, alongside their other effects, the easiest possibility would be leakage if a crack/corrosion grew to some extent to rupture a pipe. Hence, all the faults should immediately be identified once they occur.

Acoustic emission had long ago been utilized as the effective means to inspect materials and structures. This phenomenon occurs within a material from localized sources once the material is put stress and strain on. Acoustic emission is extremely sensitive to structural irregularities; therefore, AE signal–based techniques have been employing in recent studies of pipeline fault diagnostics. However, because the content of AE signals is extraordinarily complicated, which is resulting from the effects of wave attenuation and dispersion as well as the interference of ambient noises, the algorithms that scientists and researchers developed may not be reliable enough to extract accurate information about a defect from the signals. In this dissertation, what need to be improved in the pipeline fault diagnosis using AE signals will be considered, in which the identification of leak and crack is concentrated.

First, the thesis introduces a reliable method to detect a leak in a water pipeline based on AE wave attenuation. This approach segmented AE signals into short frames, calculated intermediate quantities that contain the symptoms of a leak and keeps its characteristic adequately stable even when the environmental conditions change, and trained a k–NN classifier using features extracted from the transformed signals to identify a leak in the pipeline. Experiments were conducted under different conditions to confirm the effectiveness of that method. The experimental results indicate that it offers better quality and more reliability than using features extracted directly from the AE signals to train the k–NN classifier. Moreover, the proposed method required less training data than existing techniques. The transformation method was highly accurate and worked well even when only a small amount of data was used to train the classifier, whereas the direct AE–based method returned misclassifications in some cases. In addition, the robustness of the proposed method was also tested by adding Gaussian noise to the AE signals. The proposed method was more resistant to noise than the direct AE–based method.

Leak localization is as important as leak detection in the general pipeline fault diagnosis. As a result, this dissertation proposes a method of leak localization using AE bursts, which is aimed particularly at industrial–fluid pipelines made from steel. The proposed approach exploited a burst phenomenon in AE signals and combined signal processing with a physical wave–propagation model in order to improve the leak localization. The algorithm sought AE bursts based on a detection theorem and then associated neighboring bursts into unique burst groups. Filtered AE events which are pairs of grouped bursts from two signal channels in turn allowed the extraction of precise location where a leak came about. The resulting localization method yielded a mean error of approximately 2.5% of the distance between two sensors, while this parameter returned by conventional approaches was greater than 10%. The combination of grouping and filtering in the methodology elucidates event concentration and reduces error.

Another technique is also offered for detecting a leak in a gas pipeline using a *k*–NN classifier and hybrid AE features in the thesis. This whole algorithm was embedded in an MCU to achieve a complete real–time pipeline leak detection system. First, AE signals were first recorded from a gas pipeline testbed under various conditions and offline investigated to synthesize the leak detection algorithm. The approach explored different features of normal/leaking states from corresponding datasets and eliminated inferior features to enhance the performance of leak detection. In order to obtain the robustness, alongside features were normalized, the trained *k*–NN classifier was adapted to the specific environment where the system was installed. Furthermore, the system decided the state of the pipeline on ALEOR and a defined threshold to reduce false alarms. The entire proposed system was implemented on the 32F746G–DISCOVERY board, and to verify this system, numerous real AE signals stored in a hard drive were transferred to the board. In the experiments, the implemented system executed the leak detection algorithm in a period shorter than the total input data time, thus guaranteeing the real–time characteristic. Additionally, the system yielded high ACA despite adding a white noise to input signal, and false alarms did not occur with a reasonable ALEOR threshold.

Finally, a novel approach is presented in the dissertation to detect and localize a crack in a pipeline transporting fluid under high pressure using AE signals. From signals acquired by two R15i-AST sensors at two ends of a fluid pipeline, the proposed method scanned peaks in the individual signal channels in the time–frequency domain and filtered out noise to obtain AE events. Subsequently, adjacent events were combined into grouped events, and these were picked and paired together on two sensor channels to localize emission sources using the TDOA technique. To improve the location accuracy, the mechanism only determined TOA of Rayleigh waves with a similar frequency in event pairs. Furthermore, the Rayleigh wave velocity was

found by a PLB procedure. Additionally, false emission sources were eliminated by considering the wave energy attenuation characteristics in their propagation path. After locating the emission sources, the approach observed their distribution according to the position and time of occurrence. The variation in AE activity against applied load, which was established by counting the returned sources, could indicate irregular structural changes in a material. The location of the structural change could be surmised by the emission source distribution and density according to the position along the pipeline. Experimental results showed that the proposed method correctly diagnosed faults in the considered pipeline from AE signals, whereas a conventional approach (performed by detecting hits with a threshold) inaccurately localized AE sources and imprecisely exposed signs of abnormal structural transformation

Contents

List of Figures

List of Tables

Nomenclature

ABM	Acoustic emission burst monitoring
ACA	Average classification accuracy
ACC	Ambiguity correlation classification
ADC	Analog–to–digital converter
AE	Acoustic emission
ALEOR	Accumulative leaking event occurrence rate
ANN	Artificial neural network
AVA	Average amplitude
CC	Cross–correlation
CCF	Cross–correlation function
CCTV	Closed Circuit Television
DAQ	Data acquisition
DWT	Discrete Wavelet transform
EMD	Empirical mode decomposition
ETY	Entropy
FFT	Fast Fourier transform
GCC	Generalized cross–correlation
kHz	Kilohertz
KUS	Kurtosis
KL	Kullback–Leibler
k–NN	k nearest neighbors
MHz	Megahertz
LMD	Local mean decomposition
MCU	Microcontroller unit
ME	Mean
NI	National Instruments
PCA	Principal component analysis
PHAT	Phase transform

PLB	Pencil lead breaking
RMS	Root mean square
ROC	receiver operating characteristic
SCT	Smoothed coherence transform
SKE	Skewness
SPC	Spectral centroid
SPP	Spectral peak
SPS	Spectral spread
STD	Standard deviation
STE	Short time energy
SVM	Support vector machine
TOA	Time of arrival
TDOA	Time difference of arrival
US	United States
VMD	Variational mode decomposition
WDD	Wavelet decomposition and denoising
WPT	Wavelet packet transform
ZCR	Zero crossing rate
2–D	Two dimensional
3–D	Three dimensional

Chapter 1

Introduction

1.1. Motivation

1.1.1. Urgency of Pipeline Leak Detection

Nowadays pipelines play a vital role in human life and industry. First, people utilize them to conduct clean water to buildings, hospitals, schools, etc. where inhabitants almost lean on such water supplies. Second, to start vehicles on roads, to operate machines in factories, to warm houses in winter, or to cook meals in kitchens and restaurants, humans need fuel, for example, oil and gas, and pipelines are used for transporting fuel from mining and refinement places to stations. Also, other industrial liquids such as benzene, xylene, and toluene are widely used in industry, and they are transported via pipelines. Moreover, domestic wastewater is discharged from houses, markets, etc. to proper treatment plan through pipelines.

Figure 1. 1. Major fire after a pipeline oil leak in Cairo, Egypt (source:
*https://www.energyfacts.eu/shuqair-mostorod-pipeline-oil-leak-causes-major-fire-break-out-in-cairo/***)**

Any leak in a pipeline system can lead to serious consequences. For example, a leak occurring in a pipeline network transporting clean water, gas, oil, or other industrial liquids, which are extremely valuable resources, can cause waste of resources and then make significant loss of finance. The total cost due to such a failure can amount to a few million US dollars in some case [1]. A pipeline leak can also pollute the environment and affects human health. For instance, oil from an undersea leaky oil pipeline can spread on the sea and destroy the nearby ecosystem where billions of marine creatures live. Then, fishermen accidentally catch fishes at that contaminated location and bring them to markets. Those fishes may be incredibly poisonous to consumers. Furthermore, a leak in a natural gas pipeline in a crowded place would directly risk

1

human injury and death because natural gas could lead to violent fires and explosions. Additionally, wastewater would decay into flammable and toxic substances; therefore, it could be terribly dangerous if there was a leak in a wastewater disposal pipeline system. Figure 1.1 is a concrete example of a leak oil which caused a major fire in Cairo, Egypt suburb on July 14th, 2020, and injured many people.

Figure 1. 2. Uncontrolled pipeline leaks (sources: (a) *https://www.torque-lock.com/dont-ignore-water-damage-warnings/,* (b) *https://www.siliconrepublic.com/machines/mit-pipeguard-water-leaks,* (c) *https://www.cfr.org/blog/trouble-oil-pipelines-nigeria*)

Figure 1. 3. Example of corroded pipeline (source: *https://www.petroskills.com/blog/entry/00_totm/aug20-sub-oilfieldcorrosion#.YV0-qppByUk*)

Figure 1. 2 shows several uncontrolled pipeline leaks. Such leaks might be inevitable even though the pipeline is always designed and installed strictly and properly according to standard technical principles and criteria as introduced in [2, 3] to ensure the safety in the operation of pipeline. First, an unexpected leak may be a result of assembly defects such as loosing of screw thread connections. Furthermore, due to corrosions and material aging, pipelines can gradually be degraded, and they can break in a long period of operation. An example of a pipeline which

has seriously been corroded is shown in Figure 1. 3. Another inherent risk is that gas or liquid flowing through a pipeline may trigger off twinning, boundary sliding, and moving dislocations in materials that the pipeline is made from and then stimulate crack growth within the pipeline and rupture it anytime as illustrated in Figure 1. 4. Moreover, external forces including ground movement, nearby strong vibrations, etc. can damage pipelines and create leaks. The 11th report of the European Gas Pipeline Incident Data Group [4] concludes that there are 1,411 pipeline incidents that are recorded in the period from 1970 – 2019 on 142,711 km of pipelines and their primary causes are external activities (e.g. digging, piling, ground works), external equipment (e.g. anchor, bulldozer, excavator), corrosion, construction defects, material failure, design error, lightning, maintenance error, erosion, landslide, and flood.

Figure 1. 4. Fatigue crack of a pipeline (source: https://https://www.ishn.com/articles/108930-fatigue-crack-cause-of-pipeline-rupture-oil-spill)

1.1.2. Existing Technologies for Pipeline Fault Diagnosis

To reduce the dangers posed by a pipeline leak, people try their hardest to inspect regularly pipelines and early detect any leak there; thus, the necessary maintenance and repair as well can be conducted immediately if a leak happens. However, usually, a pipeline network is highly sophisticated and some of its pipe segments can be buried deep underground to link many places together. As a result, pipeline operators would be difficult to discover a leak immediately without an automatic leak detection method. Researchers have proposed various techniques that can spot a leak in a pipeline automatically as reviewed by Murvay et al. [5], Leman et al. [6], Datta et al. [7], Chan et al. [8], Moubayed et al. [9], and Baroudi et al. [10].

Figure 1. 5 illustrates current detection technologies referred to the categorization sketched by Chan et al. [8]. Vision utilization, in terms of the passive leak detection, is to observe anomalies

on ground surface or vegetation growth, i.e. surrounding environments, hence spotting a leaky place. This is the oldest leak detection solution and mainly carried by human power. As stated above, this method would be inapplicable for the underground pipeline networks where are not easy to observe or the leak detection cannot be early enough because it may take a long time to manifest the leak syndromes on the ground surface and to affect the nearby ecosystem. A more advanced approach is sensor utilization carried by using measurement devices. This solution not only can improve the leak detection accuracy in a pipeline but also be applicable for the large pipeline systems. For example, the CCTV system consisting of a remote–controlled pan and a camera is installed on a robot inspecting inside pipes. This technology can precisely discover any abnormality of the pipelines.
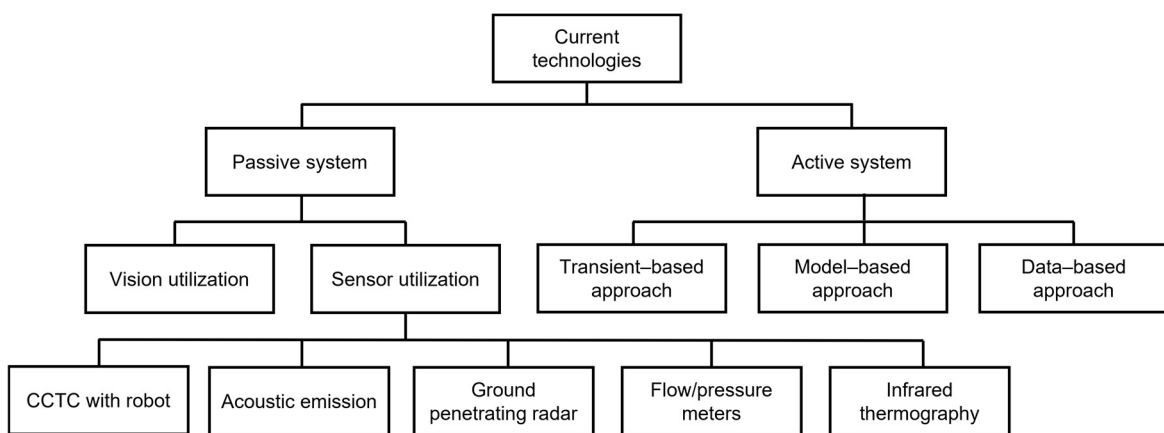
Figure 1. 5. Categorization of current leak detection technologies

Other approaches involve acoustic emission mechanisms, ground penetrating radar, flow/pressure meters, and infrared thermography using. AE techniques are constructed on the phenomenon of release of elastic waves from an active source [11]. AE signals coming from a leak through which liquid/gas passes can be captured by AE sensors installed on the surface of pipes due to the wave propagation along the pipeline; thus, the leak can be detected and localized. However, AE signals depend on various factors such as pipe material, operating conditions. They are also subjected to ambient noises, while the working environments of pipelines are extremely noisy. The operation of a ground penetrating radar used for leak detection is based on analyzing the electromagnetic signatures of an image of the pipe underground. Leaky regions would manifest their discrimination in features in the captured images. A leak can then be localized using the direct interpretation of the images. Yet the reflection of electromagnetic wave is strongly influenced by the characteristics of soil where the pipelines are installed. For instance, the moisture of soil can absorb the electromagnetic energy; thus, the resulting signals may not be sufficient to establish a correct portrait of the pipelines beneath. The infrared thermography method utilizes the infrared radiation across

spatial surfaces and energy transfer theorem detect a leak in a pipeline based on the resulting temperature difference in the vicinity of the leak. This approach depends on weather and surface conditions as well as soil types. A pipeline leak can be also detected using flow rate or pressure meters and observing a mass–balance operation. However, these meters may not be sensitive to subtle variations in flow rate and pressure caused by tiny leaks.

Besides, the solutions belonging to the active systems such as transient–based [12], model–based [13], data–driven based [14] approaches are proposed. In nature, these methods are the implementation of sensor utilization, in which they employ signals collected by the corresponding sensors as their input measurements. Commonly, the transient–based methods indicate the presence of leak using transient negative pressure waves measured within a pipeline system. They analyze the data collected during the occurrence of transitory events via minimization of the difference between the observed and the calculated parameters. These methods are subjected to background noise or other events happening in a complex network. Negative pressure waves also depend on geometric parameters of pipe and fluid properties. The model–based approaches usually exploit mathematical functions to express the operation of a pipe system. It can detect a leak and determine its approximate location using pressure measurements and their estimation resulting from a hydraulic network model. However, to achieve a high accuracy, the model should be sufficient to represent the network. The data–driven method detects a leak relying on the signal processing and statistical analysis using the data collection. This approach is advantageous because it does not require any specific in-depth knowledge about the inspected object but learns from the historical data.

1.1.3. AE Signals–based Pipeline Fault Diagnostics and Thesis Objectives

Among the existing technologies, the leak detection using AE signals is promising for several reasons. First, AE sensors can offer high sensitivity to any anomaly within materials. Therefore, the AE signal–based method is potential for immediately identifying a small–scale leak even crack growth which could lead to an imminent aperture in a pipeline. Second, the method extracts information about pipeline health from signals collected by AE sensors attached to the outer surface of pipe (see Figure 1. 6), which does not disturb any operation of the pipeline system. Thus, it belongs to the class of non–destructive and non–invasive techniques which are used to test a part or material or system without impairing its future usefulness. Moreover, AE testing is nondirectional. The energy that is detected in AE testing is released from within the test object rather than being supplied by the test method. Most AE sources radiate energy in spherical wave fronts and a transducer located anywhere in the vicinity of an AE source can

sense the resulting AE signals. This contrasts with other methods of nondestructive testing, which rely on prior knowledge of the probable location and orientation of a discontinuity to direct an energy beam on a path that will properly intersect the area of interest, for instance, ground penetrating radar and ultrasonic scanning. Thus, the method using AE signals for leak detection in a pipeline has attracted a lot of interest [15-19].
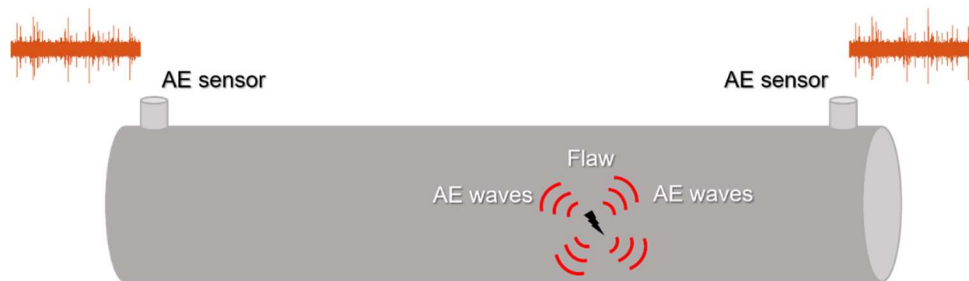
Figure 1. 6. AE signal–based pipeline fault diagnosis

Formally, AE is a phenomenon in which transient elastic waves are emitted by the rapid release of energy from localized sources in a material, or transient elastic waves so generated [11]. Hence, diverse AE sources could be constituted by the operation of pipeline such as mechanical deformation of stressed materials, leaks, cavitation, and friction. Additionally, elastic waves suffer from dispersion and attenuation during their propagation through a pipeline [11, 20, 21]. This would severely distort signals before AE sensors can acquire them. Background noise also has a considerable impact on the AE testing. Noise can come from nearby operating machines, vehicles, and environmental agents such as rain and wind; or it is created by flow turbulence inside the pipeline. In consequence, usually, it is not straightforward to obtain correct information about health of an operating pipeline from AE signals because this information is deeply buried in the content of signals.

To determine whether AE sources are caused by internal failures in a pipeline or are irrelevant sources and to localize them, researchers have been proposed a series of different methods using advanced mechanisms of signal processing and pattern recognition. Sun et al. [22, 23] employed LMD and WPT to get the most intrinsic components from AE signals collected in a natural gas pipeline, then identified leak apertures by envelope spectrum entropy or RMS entropy features and an SVM classifier; finally it analyzed the time–frequency parameters to obtain time delays for leak localization. Chencheng et al. introduced an adaptive noise cancelling method based on EMD, combined with the CC between two sensor channels to eliminate poorly correlated elements that are less related to a leak, thus increasing the signal–to–noise ratio of leak signals and improving the error leakage location [24]. Gao et al. established a procedure to improve the

shape of the cross–correlation function for leak detection in a plastic water distribution pipeline, hence resulting in unambiguous and clear estimate of the time delay [25]. Qiyang et al. also suggested another adaptive denoising approach using VMD and ACC to detect a small leak in a natural gas pipeline [26]. Zhenlin et al. proposed using kernel PCA to reduce feature dimensions and select optimal features, and an SVM classifier to recognize leakage levels in a gas pipeline valve [27]. Yu et al. offered a method detecting a leak at screw thread connection for indoor gas distribution using time– and frequency–domain features directly extracted from raw AE signals, selecting the best features relying on the distance from the noise background, and training an SVM classifier for leak detection [28]. Suzhen et al. also extracted features from AE signals in time and frequency domains, but selected the best signatures based on KL distance, and finally conducted them to an ANN to identify leakage of a socket joint between two pipeline segments [29]. Nicola et al. developed a system in LabView programming environment for leak detection and localization in a pipeline using the CC method and a data acquisition system based on AE sensors [30]. Song et al. suggested selecting features by analyzing average and CC of the features to remove redundant signatures, and applied an artificial neural network model for the leakage classification in galvanized steel pipes due to loosening of screw thread connections [31]. Rui et al. used the DWT with the Wavelet basis selected by the entropy–based algorithm to denoise signals and then extracted features and chose features using the Relief–F mechanism, and trained an SVM classifier for leakage severity identification in a gas pipeline [17]. Xu et al. proposed denoising AE signals using an improved VMD, optimizing the governing parameters via coupling particle swarm optimization to a fitness function based on maximum entropy, and then extracting features from reconstructed signals based on the energy ratio of each VMD sub–mode and training an SVM classifier for leakage pattern recognition in a liquid pipeline [32].

Most researchers have been adopting data–driven techniques to detect a leak in a pipeline from AE signals. In those, classifiers are trained by previously recorded datasets to categorize the states of pipeline health. A data–driven model is more convenient than another one built directly from mathematical equations because AE signals constantly fluctuate. However, such classifiers would be insufficient in diverse cases due to lack of data for training. On the other hand, the time–delay methods the researchers have been exploiting for leak localization would not result in high accuracy because they utilized the CC function between two signals while the CC is rather low. As a result, this dissertation further digs deep into the AE signal–based pipeline leak identification. The problem of pipeline crack detection and localization is also necessary and urgent [33, 34] because it can help prevent following leaks. Thus, the objective

of the thesis focuses on pipeline leak and crack identification using AE signals, specifically, including: (i) improving reliability and robustness of leak detection, (ii) increasing accuracy of leak localization, (iii) crack detection and localization.

1.2. Dissertation Outline

The thesis is composed of six chapters, and they are briefly outlined as follows.

Chapter 1 states the motivation and the outline of thesis.

Chapter 2 introduces a reliable method for leak detection in a water pipeline based on AE wave attenuation. This approach transformed AE signal segments into intermediate quantities to identify a small leak using a k–NN classifier. *This work was published in the international journal, Energies, in 2019.*

Chapter 3 proposes a leak localization method for an industrial–fluid pipeline using AE bursts. The proposed approach combined signal processing, which was used for AE burst detection, with a physical wave–propagation model to locate a leak in a steel pipeline. *This work was published in the international journal, Measurement, in 2020.*

Chapter 4 offers a technique using a k–NN classifier and hybrid features extracted from AE signals for detecting a gas pipeline leak. The scheme of detection algorithm was implemented in an MCU to achieve a complete real–time leak detection system. *This work was published in the international journal, Sensors, in 2021.*

Chapter 5 presents a novel approach to crack detection and localization in a fluid pipeline using AE signals. The proposed method found events in the signals in the time–frequency domain and then localized them based on the TDOA technique. To reduce the location error, the algorithm only utilized events belonging to the wave mode Rayleigh with the same frequency in event pairs from the two signal channels. *This work was published in the international journal, Mechanical Systems and Signal Processing, in 2021.*

Finally, Chapter 6 provides a summary of the contributions and a discussion of future work.

Chapter 2

Water Pipeline Leak Detection utilizing Acoustic Emission Wave Attenuation

2.1. Introduction

It is difficult to establish a detection model directly from mathematical equations due to the complication of AE phenomenon; therefore, numerous methods training the model using recorded datasets have been proposed [29, 35]. Although these techniques offer high accuracy, they could be inefficient in diverse circumstances because the classifiers are trained using features extracted directly from AE signals. Thus, it makes the classifiers dependent on the absolute signal levels that are influenced by the flow rate and pressure in a pipeline system [3, 21]. Moreover, AE signals also vary with temperature, which is uncertain in industries. As a result, AE measurements recorded under such uncertain working conditions have different signal levels, and thus, a fault diagnosis model based on the absolute amplitude values of the signals is unreliable. In [36], the authors presented a study that investigates the inherent properties of the vibro-acoustic signals instead of relative amplitude values. It delivers useful insights and results for water leak detection, but the study focused on pipelines buried under soil less influenced by external factors while AE signals from pipelines above soil in factories are subjected to noise. This issue can be addressed by acquiring many training datasets under different working conditions to provide enough information for the classifier. However, it is not an optimal solution because the design of experiments can become immensely expensive or even prohibitively complicated. Therefore, this study extracts features from transformed signals independent of an absolute level instead of directly from AE signals. The obtained classifier can detect small leaks in a pipeline with high accuracy and reliability.

For the experiments, this study uses a water pipeline system deployed in a laboratory. Three AE sensors are mounted on different parts of the system with the known distances among the sensors. A pinhole is drilled through the pipe wall to simulate leakage of the pipeline and a valve is attached to the wall to control the flow of water through the leak. When the valve is closed, and system is working in a normal state (no leakage) then there is no variation in the levels of the recoded signals. If a small leak appears between sensors 2 and 3, an imbalance is created between AE channels 1 and 2 by reporting dissimilar amounts of noise from the leak. Since sensor 2 is nearer the leak than sensor 1, the leak noise on channel 2 is greater than that on channel 1, as explained by the theory of wave attenuation [37-39]. Thus, this uneven behavior of the signals can be used to detect small leaks in a pipeline system. The leak-signal-

to-normal-noise ratio and the attenuation law of wave propagation can be used to detect the leak; these characteristics are demonstrated in the methodology section.

To demonstrate the advantages of the balance/imbalance–based approach, this paper uses the theory of KL distance [40] to measure the separability between the two classes (i.e., NORMAL and ABNORMAL) in different experiments. Furthermore, it applies a *k*–NN model to identify a normal/leak state. Two approaches, one using features extracted directly from the AE signals and the other using features extracted from the transformed signals based on the function $g(r)$ are employed to train *k*–NN classifiers. These classifiers are trained on the data of one working condition and tested using data of other conditions. Additionally, Gaussian noise is added to the AE measurements to evaluate the robustness of the classifiers.

2.2. Data Acquisition

2.2.1. Testbed Configuration
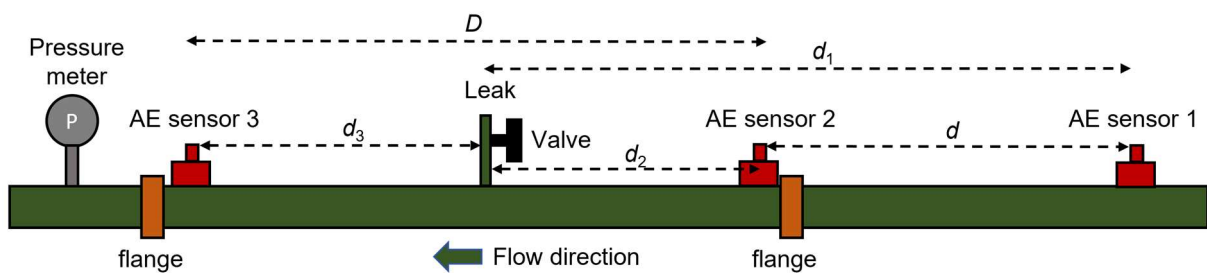
Figure 2. 1. Data acquisition setup

Figure 2. 2. Pipeline testbed

Figure 2.1 and Figure 2.2 show a setup of the acquisition of AE signals from a water pipeline system. The configuration parameters of the testbed are listed in Table 2.1. In the experiments, the pipeline leaks were emulated by four orifices with different diameters including 2.0 mm,

10

1.0 mm, 0.5 mm, and 0.3 mm, which are abbreviated to L1, L2, L3, and L4, respectively. The water flow was controlled using pressures of 7, 13, and 18 bar, given as P1, P2, and P3, respectively. The experiments were conducted under a stable temperature of approximately 29 °C. To acquire AE signals, R15i–AST sensors were used to provide high sensitivity during the data recording and the recorded signals tend to be free of low-frequency components. Their characteristics have been detailed in [41].

Table 2. 1. Experimental parameters

No.	Quantity	Detail
1	Distance from Sensor 2 to Sensor 3 (D)	2000 [mm]
2	Distance from Sensor 2 to Sensor 1 (d)	1000 [mm]
3	Distance from Sensor 2 to Leak (d_2)	900 [mm]
4	Distance from Sensor 3 to Leak (d_3)	1100 [mm]
5	Thickness of pipeline	6.02 [mm]
6	Outer diameter of pipeline	114.3 [mm]
7	Material of pipeline	Stainless steel 304

2.2.2. Data Record

The normal and abnormal states of the system refer to the closed and open positions of the valve installed on the leak, respectively. The data for every pair of (P, L) combinations are recorded for 2 minutes with a sampling frequency of 1 MHz after the water flow are stable. Thus, there are a total of 72 datasets for the three signal channels in the experimental conditions (3 channels × 3 pressures × 4 leaks × 2 classes). Figure 2.3 presents mode setting of water flow. In the first stage, the pump is turned on, the leak is deactivated (closed) and the pressure is controlled around 7 Bar. Then, AE signals are recorded in 2 minutes. In the second stage, the leak is activated (open). The acquisition device waits for the flow stabilization and records AE signals in 2 minutes. The process continues, as shown in Figure 2.3.
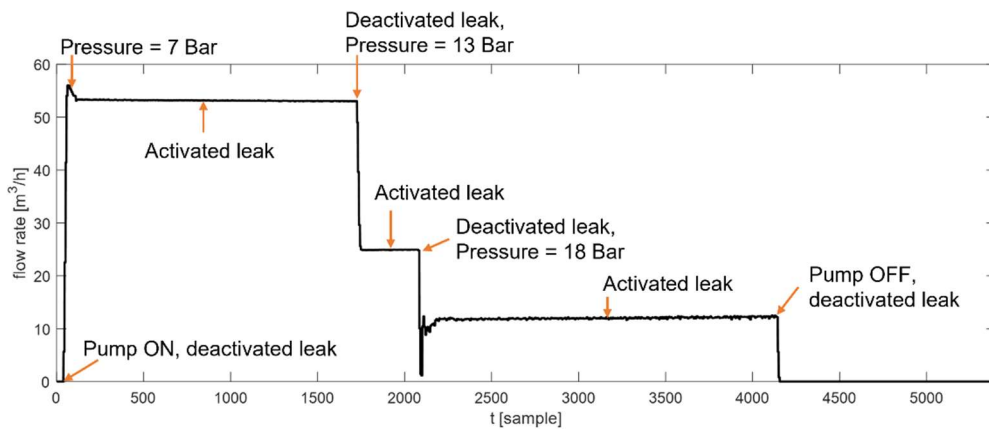
Figure 2. 3. Input flow rate experimented with leak L3

2.3. Small Leak Detection Methodology

The mathematical modeling of the leaks and their symptoms is based on the testbed, as shown in Figure 2.1.

2.3.1. Symptoms of Leak Presence

There exist AE activity in pipelines, even when they are operating in a healthy state. It can be due to the mechanical sources such as pumps or particles in the flow hitting a pipeline wall or hydraulic sources caused by pressure pulses at vortexes in the fluid inside the pipeline [3]. Let assume that n_i (*i =1,2*) are AE signals acquired by sensors 1,2 in the normal state, and their mean and variance are 0 and $N_1 \approx N_2 \approx N > 0$, respectively. When a small leak occurs in the testbed pipeline, it makes a small disturbance in the flow around the leak, which introduces a new AE source into the system. The previous source n_i can be deemed as background noise; suppose that it is uncorrelated with the source of the signal obtained from the leak. Therefore, the model for AE measurements from the sensors for this scenario can be given as: $z_i = s_i + n_i$ (*i =1,2*); where s_i is the leak AE signal received by sensor *i*. If the variance of z_i and s_i is Z_i and S_i, respectively, then the uncorrelation between n_i and s_i, $Z_i = S_i + N$ can be explored by setting $g = Z_2/Z_1$ and it is transformed as follows:

$$g = \frac{Z_2}{Z_1} = \frac{S_2 + N}{S_1 + N} = \frac{S_2/N + 1}{S_1/N + 1} \tag{2.1}$$

If the background noise n_i (*i =1,2*) is bandlimited white noise, then its variance is always N over its entire frequency range. Next, consider the measurement model when the measurement consists of the background noise n_i and the leak signal s_i (*i =1,2*) at a frequency ω. In fact, the spectrum of the leak noise is a broadband range of frequencies; however, all the components demonstrate an identical behavior toward the leak phenomenon. The AE attenuation characteristic from a power law [39], S_1 and S_2 are related by:

$$S_1 = S_2 e^{-\alpha d} \tag{2.2}$$

where $\alpha = \alpha(\omega)$ is the attenuation coefficient in wave propagation, which is dependent on frequency, *d* is the distance between sensors 1 and 2, and S_i (*i =1,2*) is the leak signal having frequency ω. By substituting (2.2) into (2.1), and abbreviating $r = S_2/N$, which is the SNR measured by sensor 2 at frequency ω, and symbolizing $\beta = e^{-\alpha d}$, (2.1) becomes:

$$g = \frac{r+1}{\beta r + 1} \tag{2.3}$$

Next, take the partial derivative of g according to r:

$$\frac{\partial g}{\partial r} = \frac{1-\beta}{(\beta r + 1)^2} \tag{2.4}$$

Since $0 < \beta < 1$ for $0 < d < \infty$, then $\partial g / \partial r > 0$. As a result, $g(r)$ is a monotonically increasing function according to r. If $r_1 \neq r_2$, then $g(r_1) \neq g(r_2)$. Naturally, a normal state has $r = 0$ at every frequency ω, and an abnormal state always has $r \neq 0$; thus, the function $g(r)$ is applicable for leak detection. Figure 2.4 shows the dependence of $g(r)$ on r with different β values at a particular frequency. It can be easily observed that all the curves $g(r)$ increases from the normal state when r increases.
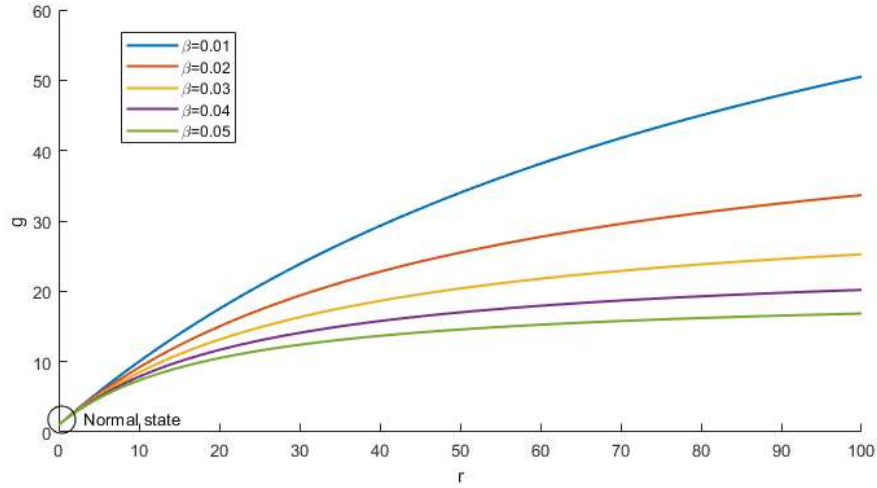
Figure 2. 4. The dependence of $g(r)$ on r with different β at frequency ω

2.3.2. Robustness of g(r) in Leak Manifestation

This section investigates the reliability of leak manifestation using the function $g(r)$ when the background noise increases. Gaussian noise is added to the signals to emulate the presence of noise. Suppose that an amount of noise Δn with the mean 0 and variance ΔN is added to the noise background while the leak signal remains the same. At this moment, the background noise n is replaced by $n' = n + \Delta n$; its variance is $N' = N + \Delta N$, and r is replaced by $r' = S_2/N'$.

Setting $\gamma = \Delta N / N$, produces $r' = r/(1+\gamma)$, which is a function of two variables (r, γ). Consider the partial derivative of r' with respect to γ:

$$\frac{\partial r'}{\partial \gamma} = -\frac{r}{(1+\gamma)^2} \tag{2.5}$$

Now, the function g is replaced by (2. 6):

$$g^{'} = g\left(r^{'}\right) = \frac{r+1+\gamma}{\beta r+1+\gamma} \tag{2. 6}$$

and its partial derivative according to γ is given by:

$$\frac{\partial g^{'}}{\partial \gamma} = \frac{\left(\beta-1\right)r}{\left(\beta r+1+\gamma\right)^{2}} \tag{2. 7}$$

If $0 < \beta < 1$ and $\partial g^{'}/\partial \gamma < 0$, then $g^{'}$ decreases when γ increases. In other words, the discrimination quality of $g^{'}$ becomes poorer as the intensity of the background noise is higher. This characteristic is similar to $r^{'}$; however, $g^{'}$ is more reliable than $r^{'}$ because the decline in $g^{'}$ is smaller than the decline in $r^{'}$. Also, we obtain:

$$\frac{\Delta_2}{\Delta_1} = \left(1-\beta\right)\frac{\left(1+\gamma\right)^{2}}{\left(\beta r+1+\gamma\right)^{2}} \tag{2. 8}$$

where $\Delta_1 = \left|\dfrac{\partial r^{'}}{\partial \gamma}\right|$, $\Delta_2 = \left|\dfrac{\partial g^{'}}{\partial \gamma}\right|$.

Obviously, if the parameter β in (2.8) is selected suitably, then $\Delta_2/\Delta_1 << 1 \ \forall r, \gamma$. As a result, $g(r^{'})$ varies more slowly than $r^{'}$ if γ is increasing. It turns out that if the noise background increases to some extent, the variable $r^{'}$ exceeds the limitation of leak discrimination, whereas the function $g(r^{'})$ still provides enough differentiation.

In (2.8), if β approaches 0, the ratio Δ_2/Δ_1 converges 1 and the variation of the function g is similar to r if the background noise changes, the characteristic of g is no longer robust. In contrast, (2.3) reveals that if β approaches 1, g converges 1 for any r. This means that the function g does not manifest any abnormal state of the system. Hence, the parameter β should be chosen optimally to trade off between the two above cases so that both high sensitivity and reliability can be achieved.

2.3.3. Detection Procedures

This section proposes an algorithm of leak pipeline detection based on *g(r)* function because it can reliably indicate the presence of leak. Figure 2.5 shows a generic framework for leak detection using the direct AE–based and *g(r)*–based approaches. The signals from sensors 1, 2, and 3 are the inputs to the framework. In the direct method, there is no *g(r)*–construction block. After dividing the AE signals into frames, they are provided to the feature extraction block. The feature extraction process is carried out after the completion of *g(r)*–construction process.
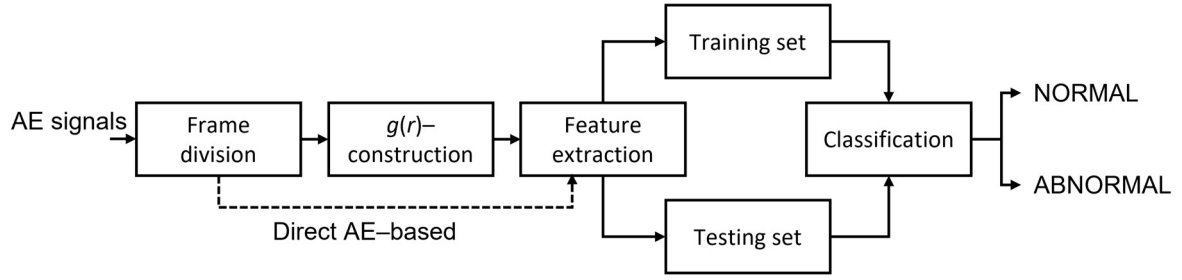
Figure 2. 5. Leak detection procedures

(1) Frame division

The recorded AE signals are segmented into a series of frames by the frame division block. The AE waves propagate through different distances from the leak to the sensors. Thus, their arrivals are lagged. It indicates that the frame indexes associated with different channels are not exactly correlation. Hence, at the detection stage, the position of the leak is obscure, and the time of arrival of the signals at the sensors is unknown. Thus, one way to deal with this problem is to select a reasonable frame size. Figure 2.6 uses an example of two signals s_1 and s_2 to explain the method. In this figure, Δt_{lag} and t_{frame} are the lag time and frame size (in time) of the two signals, respectively. Due to the existence of the lag time Δt_{lag}, a lag part of signal s_1 cannot be correlated with any part of signal s_2 in the same frame index i because it has already propagated in frame *(i - 1)* of s_2. Thus, a formula for the frame size can be defined as:

$$t_{frame} = \Delta t_{lag} + t_{ext} \tag{2.9}$$

where t_{ext} is an amount of time to extend the frame size. Obviously, the bigger t_{ext} is, the smaller the lag compared with the remains, which reduces the effect of the lag on the correlation.
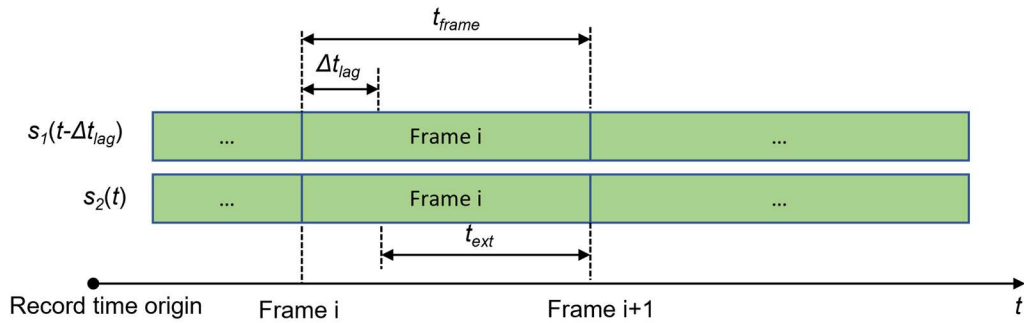
Figure 2. 6. Frame division dependent on lag time

Next, the parameter Δt_{lag} is calculated. The location of the leak in the pipeline is unknown, however, the leak lies somewhere within the tested pipeline. According to this condition, the following equations can calculate the maximum lag time and this result is used to calculate the reasonable frame size, given by (2. 10).

$$\Delta t_{lag} = \frac{D+d}{C} \tag{2. 10}$$

In (2.10) C is the wave speed. AE signals can be propagated in fluid in the frequencies range of 20 kHz to 80 kHz besides propagating through the pipe wall in higher frequencies [42]. Furthermore, leak signals are from the flow turbulence and interaction of particles at the leak point. Thus, AE signals might contain both kinds of propagation. In other words, the wave speed in water is smaller than in solid materials [43]. Hence, the value of C should be calculated with the propagation in water. The wave speed can be calculated by (2. 11) [44, 45].

$$C = \sqrt{\frac{K}{\rho\left(1+\dfrac{D_p}{e}\dfrac{K}{E}\psi\right)}},$$
$$\psi = \left[1/\left(1+e/D_p\right)\right]\left[1+2\left(e/D_p\right)\left(1+u\right)\left(1+e/D_p\right)\right] \tag{2. 11}$$

where K and ρ are the volumetric compressibility modulus and the liquid density of the medium inside the pipeline, e and D_p are the thickness and inner diameter of the pipe, ψ is a factor related to the pipe supporting condition, and u is Poisson's ratio.

The window size is counted as:

$$t_{frame} = \left(1+\zeta\right)\Delta t_{lag}, \quad \zeta = \frac{t_{ext}}{\Delta t_{lag}} \tag{2. 12}$$

where ζ must keep the lagged part of the signals, which is not very large as compared to the rest of the signal.

(2) *g(r)* - Construction

The quantity *g(r)* in Subsection 2.3.1 is formulated by dividing the variance of one frequency for the leak signal. In this section, *g(r)* vector is constituted from the signal of sensors 1 and 2. The proportions of the amplitudes over all the frequencies are considered in the *g(r)* vector (see Figure 2.7).
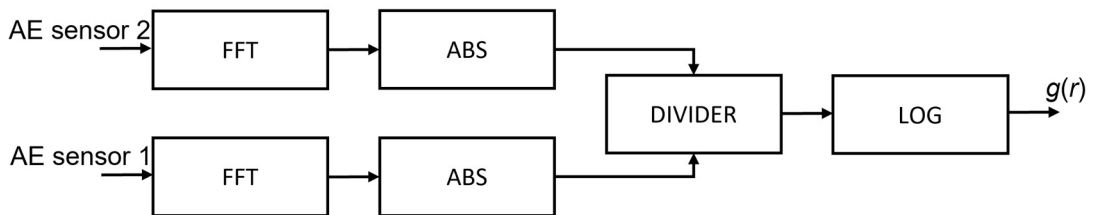
Figure 2. 7. g(r) - construction

In Figure 2.7, time domain signals are converted to the frequency domain by FFT, taking only their amplitudes as components of the divider for every frequency. After the transformation, we have a new signal in the form of $g(r)$ containing information about leakage symptoms.

(3) Feature Extraction

In this study, the three features given in Table 2. 2 are used to compare the performance of the direct AE–based method with that of the $g(r)$–based one. These features are selected because their effectiveness may be consistent in both time and frequency domains.

Table 2. 2. Typical features

No	Feature	Equation
1	Root mean square	$\sqrt{\sum_1^N x_i^2 / N}$
2	Short time energy	$\sum_1^N x_i^2$
3	Average amplitude	$\sum_1^N |x_i| / N$

(4) Classification

The k–NN kernel function is widely used to identify instances belonging to different classes during the diagnosis. The theory of k–NN is presented clearly in [40, 46]. This paper uses the k–NN classifier to solve a binary classification problem, i.e., whether a sample belong to the normal or abnormal conditions of the pipeline.

2.4. Experimental results

In this section, the proposed leak detection methodology is evaluated using the Matlab software version 2019b installed in a personal computer (CPU: Intel (R) Core i7–7800X, RAM: 16 GB, the Window 10 Pro Operating System) and the datasets of AE signals collected in the pipeline testbed as described in Section 2.2. To segment AE signals, the Hanning window function with size of 3000 points and overlap of 50% was utilized. For comparison between the direct AE–based and $g(r)$–based approaches, 6665 frames of each of patterns (P, L, 'NORMAL/ABNORMAL') and 73331 frames of those were conducted to the training and testing processes, respectively, where P ϵ (P1, P2, P3) and L ϵ (L1, L2, L3, L4) are pressures and leaks, respectively, as defined in Subsection 2.2.1. After the AE signals were divided into frames, they were transferred directly to the feature extraction block in the direct AE–based method, and to the $g(r)$–construction block in the proposed method (see Figure 2.5). Figure 2.9 illustrates AE signals of the pair (P1, L1), and Figure 2.10 presents the signals in a frame index.
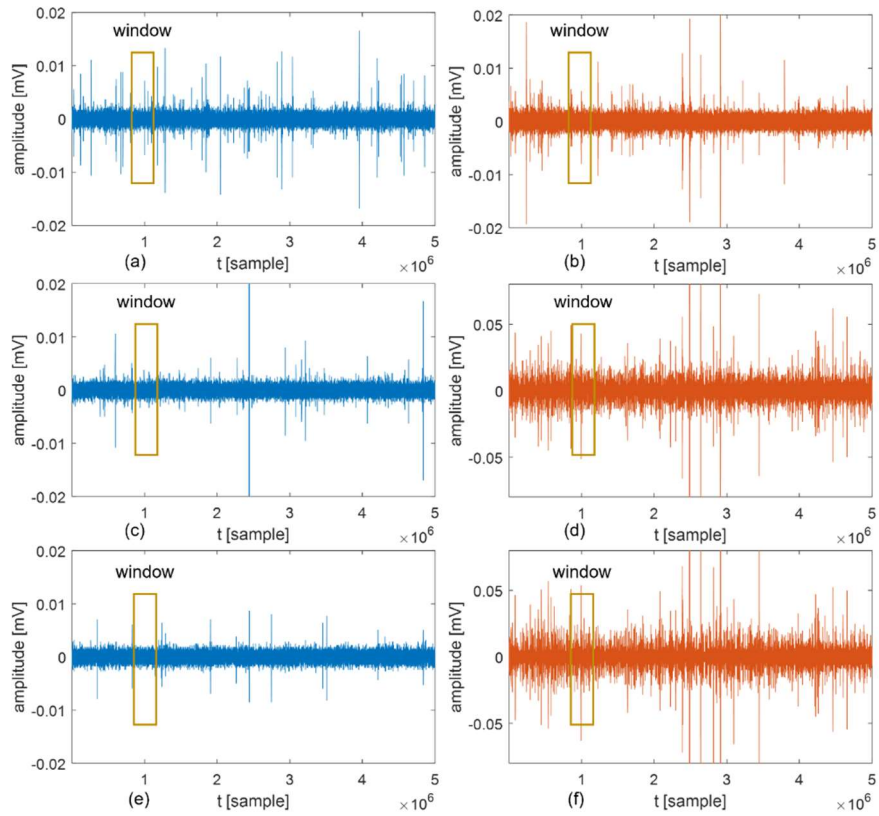
Figure 2. 8. Datasets of the pair (P1, L1): (a) Channel 1, NORMAL, (b) Channel 1, ABNORMAL, (c) Channel 2, NORMAL, (d) Channel 2, ABNORMAL, (e) Channel 3, NORMAL, (f) Channel 3, ABNORMAL
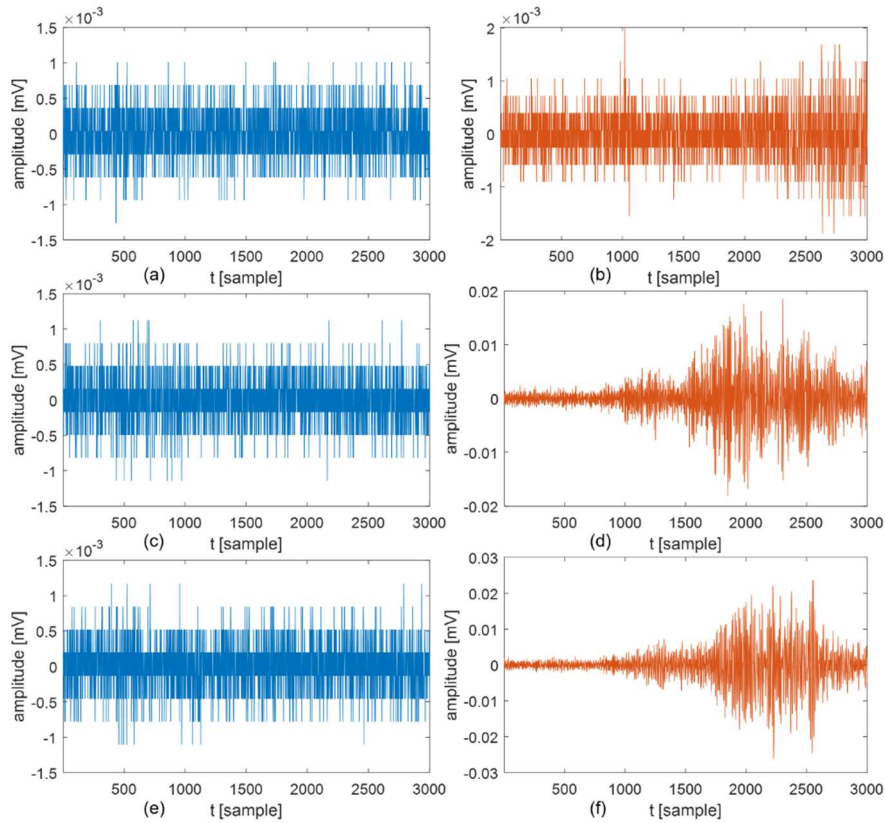
Figure 2. 9. A frame of datasets of the pair (P1, L1): (a) Channel 1, NORMAL, (b) Channel 1, ABNORMAL, (c) Channel 2, NORMAL, (d) Channel 2, ABNORMAL, (e) Channel 3, NORMAL, (f) Channel 3, ABNORMAL
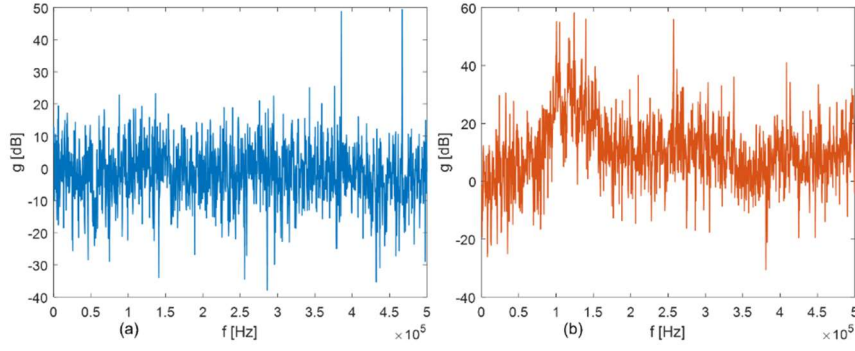
Figure 2. 10. A frame of *g(r)* for datasets of the pair (P1, L1): (a) NORMAL, (b) ABNORMAL

The signals shown in Figure 2.8 and Figure 2. 9 demonstrate that noise levels of all the channels are nearly same in the normal condition but differ from those of the abnormal condition. The channel 1 is mounted at the furthest place from the leak, the signal level at this point is the lowest which resembles to the noise while two others get higher amplitudes for being closer to the leak. Moreover, such low level of signals is prone to noise in industry as well as attenuation over a long path of wave propagation. Therefore, a classification algorithm relied on absolute levels is unreliable if it is only trained by a limited number of datasets. In a real application, the leak position is obscure, AE signals can vary from the low level like noise when AE sensors are far from the leak to higher levels when they are closer to the leak. The *g(r)*–construction block in Figure 2.7 converts the signals of channels 1 and 2 from the time domain to the frequency domain and then transforms into the quantity *g(r)*. Figure 2.10 presents a frame of *g(r)* in which the normal *g(r)* fluctuates around 0, but the abnormal *g(r)* has different trend. This behavior is independent of distance and resistant to noise. Hence, a classifier that is trained by the quantity *g(r)* can detect a leak correctly even if the environment fluctuates widely.

2.4.1. Effectiveness of the g(r)-based Approach Compared with the Direct AE–based Method

The effectiveness of the *g(r)*–based approach compared with the direct AE–based method is illustrated by state scattering in 3–D space, as shown in Figures 2.11-14. The plot uses the features and a separability comparison that relies on KL distances [40] between the two classes under various experimental conditions. The *KL* distance is given as follows:

$$d_{kl} = D_{12} + D_{21},$$

$$D_{12} = \sum p(x|w_1) \ln \frac{p(x|w_1)}{p(x|w_2)}, D_{21} = \sum p(x|w_2) \ln \frac{p(x|w_2)}{p(x|w_1)} \tag{2.14}$$

where w_1, w_2 are the two classes (i.e., NORMAL and ABNORMAL) and $x = [x_1, x_2, \ldots, x_n]^T$ is the features for n data frames, and p is the probability density function.

19
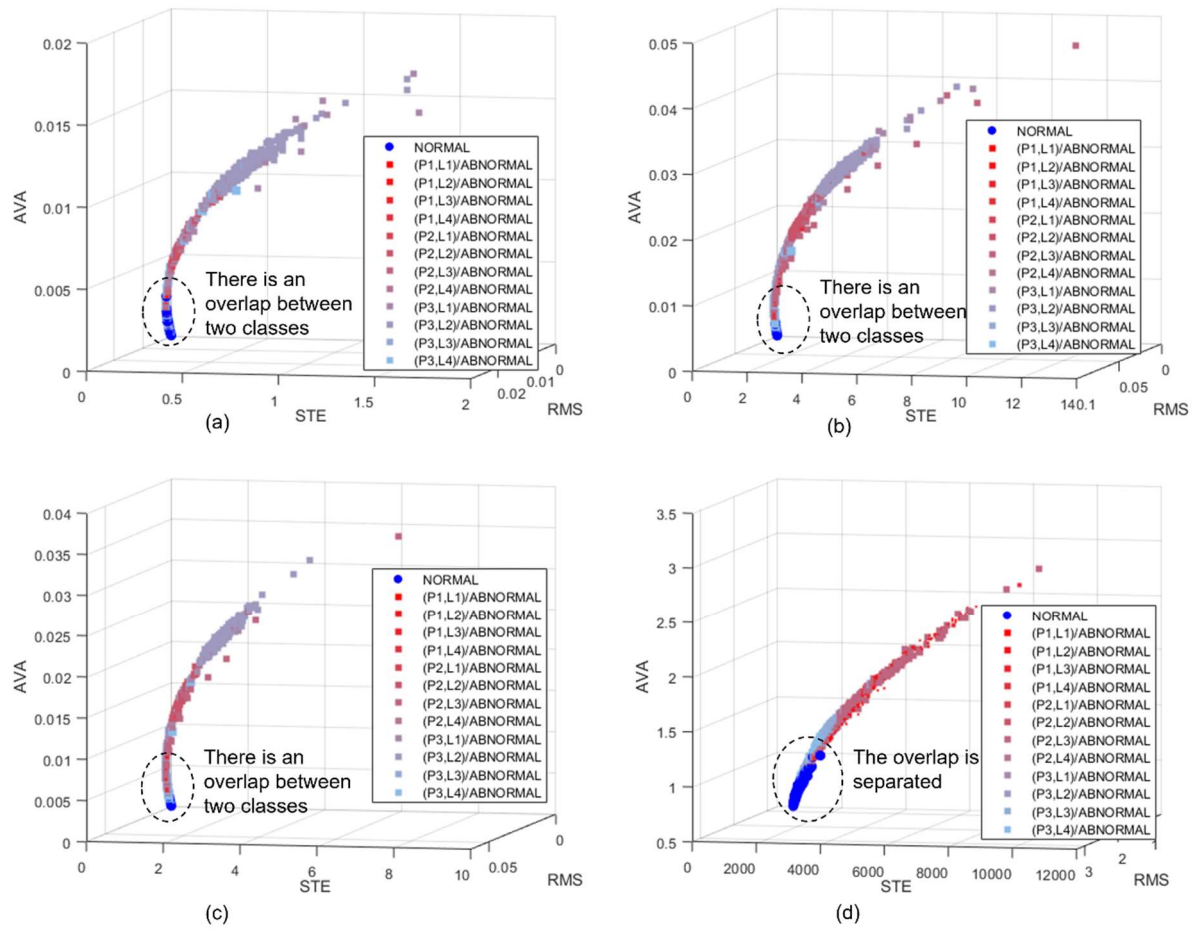
Figure 2. 11. Features–based state scattering: (a, b, and c) features extracted from signals on the individual 1ˢᵗ, 2ⁿᵈ, and 3ʳᵈ AE sensor channels (CH1, CH2, and CH3), respectively, (d) features extracted from $g(r)$

Table 2. 3. KL distances in decibels (dB) (d_{kl} (dB) = $10*\log_{10}(d_{kl})$)

P	L	CH1			CH2			CH3			$g(r)$		
		RMS	STE	AVA	RMS	STE	AVA	RMS	STE	AVA	RMS	STE	AVA
P1	**L1**	11	1	16	1	22	-10	8	25	2	59	112	60
P1	**L2**	9	-1	14	11	42	5	16	44	11	63	118	63
P1	**L3**	-6	13	2	3	34	-6	8	35	4	70	109	70
P1	**L4**	-2	6	5	11	44	6	17	46	12	63	110	63
P2	**L1**	0	24	-11	12	42	8	16	42	12	70	102	71
P2	**L2**	2	36	-7	15	58	11	21	61	17	73	104	73
P2	**L3**	3	29	-4	24	65	20	29	67	25	60	116	60
P2	**L4**	4	33	-2	13	49	8	18	51	14	68	110	68
P3	**L1**	14	48	10	23	64	19	27	65	23	72	101	73
P3	**L2**	10	48	5	17	62	12	22	65	18	76	99	77
P3	**L3**	7	-3	11	5	36	0	9	35	4	73	107	74
P3	**L4**	6	37	1	12	48	8	19	52	15	69	108	70
MEAN		5	23	3	12	47	7	18	49	13	68	108	69

Figure 2.11 shows state scattering based on the features extracted from direct AE signals and

g(r). Table 2.3 shows the *KL* distances for each pair of pressure, leak (P, L) values. Since the *KL* distances depend on features and their value can be relatively large, this study uses a logarithm calculation to convert the distances into dB for demonstration. Obviously, the *KL* distances based on features extracted directly from the AE signals vary in a wide range. In contrast, the *KL* distances calculated using features extracted from *g(r)* remain around a certain value for each type of feature under diverse conditions. Moreover, the average values also demonstrate that the class separability of the *g(r)*–based approach is greater than that of direct AE–based one. From the state scattering and the *KL* distances, it is clearly observed that the *g(r)*–based approach is more effective than direct AE–based one.

2.4.2. Classifier Training and Testing

(1) Training

This paper uses datasets of pressure and leak (P, L) pairs to train the *k*–NN classifier used in the two approaches. Table 2.4 presents the performance of the two approaches. The P4 column in the table represents the combination of datasets with all the pressures and leaks. Although Table 2.4 shows high accuracies with both methods, their performances become different when applying them under different conditions, as shown in the next section.

Table 2. 4. Trained classifiers and their accuracy (%)

	Direct AE–based				*g(r)*-based			
	P1	**P2**	**P3**	**P4**	**P1**	**P2**	**P3**	**P4**
L1	97.5	100	100		98.8	100	100	
L2	100	100	100		100	100	100	
L3	100	100	100	99	100	100	100	100
L4	100	100	100		100	100	100	

(2) Cross Testing

The previous section trained 12 classifiers using datasets for conditions corresponding to each pair (P$_i$, L$_j$) for each method: direct AE–based and *g(r)*–based. This section uses datasets in conditions (P$_m$, L$_n$) to test them, where *i, m* = 1, 2, 3 and *j, n* = 1, 2, 3, 4. The purpose of the test is to verify the reliability of each approach if the training is carried out using only some of the datasets. Tables 2.5-8 show the experimental results under this configuration. In each table, the trained classifiers are symbolized in the columns, and the test datasets are listed in the rows. It is obvious that when the testing datasets use the same conditions as the training datasets, the accuracies reach 100% or nearly 100%. However, if the testing conditions differ from the training conditions, the accuracies are not promising, and misclassification can occur when

using the direct AE–based method. In Table 2.6, an accuracy of 0% indicates misclassification. In the same situation, however, the $g(r)$–based approach identifies leaks without misclassification. Furthermore, the average accuracy of the $g(r)$–based approach is higher than that of the direct AE–based method in every test.

Table 2. 5. Accuracy (%) of classifiers trained with leak L1

		Direct AE–based			$g(r)$–based		
		P1	P2	P3	P1	P2	P3
L1	P1	98.55	69.64	62.66	99.78	99.82	99.82
	P2	99.54	99.98	65.85	99.82	99.88	99.88
	P3	96.71	99.94	100	99.88	99.82	99.8
L2	P1	99.88	100	64.45	99.98	99.76	99.76
	P2	98.23	99.86	100.0	100	99.88	99.86
	P3	99.76	100	100.0	99.98	99.96	99.96
L3	P1	100	100	62.91	99.96	99.9	99.88
	P2	99.42	99.96	67.25	100	99.82	99.82
	P3	99.32	99.98	82.05	99.76	99.82	99.84
L4	P1	99.96	64.27	62.94	99.98	99.90	99.88
	P2	99.80	99.94	63.18	99.94	99.96	99.94
	P3	99.70	99.94	63.83	99.19	99.84	99.84
MEAN		99.24	94.46	74.59	99.86	99.86	99.86
			89.43			99.86	

Table 2. 6. Accuracy (%) of classifiers trained with leak L2

		Direct AE–based			$g(r)$–based		
		P1	P2	P3	P1	P2	P3
L1	P1	65.77	62.54	0	86.92	99.54	99.9
	P2	100	63.25	62.69	62.91	99.52	99.92
	P3	100	77.53	64.53	62.65	99.58	99.92
L2	P1	100	62.65	62.51	100	99.98	99.84
	P2	99.94	100	99.36	70.4	100	99.92
	P3	100	100	100	63	99.96	99.98
L3	P1	100	62.58	62.53	98.33	99.88	99.92
	P2	99.96	64.15	62.91	98.06	100	99.92
	P3	100	62.99	62.74	98.06	99.64	99.74
L4	P1	63.97	62.58	0	64.73	99.94	99.98
	P2	99.98	62.74	62.58	98.56	99.84	99.96
	P3	99.96	63.04	62.68	95.72	98.95	99.52
MEAN		94.13	70.34	58.54	83.28	99.74	99.88
			74.34			94.3	

Table 2. 7. Accuracy (%) of classifiers trained with leak L3

		Direct AE–based			*g(r)*–based		
		P1	P2	P3	P1	P2	P3
	P1	70.51	73.06	63.19	97.96	95.1	97.83
L1	P2	99.98	99.96	76.28	81.45	66.99	79.7
	P3	99.92	99.88	100	76.79	64.07	74.84
	P1	100	100	99.82	99.98	100	99.98
L2	P2	99.86	99.80	100	99.9	99.64	99.88
	P3	100	100	100	98.95	78.53	98.64
	P1	100	100.0	63.87	99.32	99.07	99.3
L3	P2	99.96	99.96	72.89	99.98	99.7	99.96
	P3	99.98	99.96	100.	99.21	98.85	99.13
	P1	64.34	64.6	63.52	98.5	92.05	98.42
L4	P2	99.94	99.92	100	99.52	99.27	99.5
	P3	99.94	99.94	80.45	97.83	97.35	97.79
MEAN		94.54	94.76	85	95.78	90.89	95.41
			91.43			94.03	

Table 2. 8. Accuracy (%) of classifiers trained with leak L4

		Direct AE–based			*g(r)*–based		
		P1	P2	P3	P1	P2	P3
	P1	97	63.79	67.29	99.78	98.08	98.33
L1	P2	99.72	99.52	100	99.88	82.62	86.62
	P3	97.6	100	99.98	99.96	77.92	82.9
	P1	99.94	100	100	99.98	99.98	99.98
L2	P2	98.64	100	99.94	100	99.9	99.94
	P3	99.86	100	100	99.98	99.03	99.36
	P1	100	98.62	100	99.98	99.32	99.42
L3	P2	99.62	88.8	99.98	100	99.98	100
	P3	99.56	100	100	99.78	99.21	99.23
	P1	99.96	63.72	64.06	100	98.52	98.93
L4	P2	99.8	100	99.94	99.96	99.54	99.6
	P3	99.8	99.44	99.94	99.3	97.87	97.96
MEAN		99.29	92.82	94.26	99.88	96	96.86
			95.46			97.58	

(3) Testing Classifiers Trained by Combined Datasets

To detect a fault, a classifier that depends on data must be trained using many datasets representing various experimental conditions to provide high reliability. Table 2.9 shows accuracies of the two approaches when using the dataset of each pair (P, L).

Table 2. 9. Accuracy (%) of classifiers of two approaches by combined datasets

	Direct AE–based			g(r)–based		
	P1	P2	P3	P1	P2	P3
L1	96.49	99.72	97.75	99.94	100	100
L2	99.96	98.76	99.88	99.9	100	100
L3	100	99.66	99.64	100	99.92	99.96
L4	99.96	99.8	99.8	100	100	99.9
MEAN	99.1	99.48	99.27	99.96	99.98	99.97
	99.28			99.97		

The average accuracies are 99.28 and 99.97 for the direct AE–based and $g(r)$–based approaches, respectively, as shown in Table 2.9. Thus, a model trained by many datasets can work in varied conditions. However, the $g(r)$–based method is more efficient than the direct AE–based because it still achieves the expected accuracy with only a small number of training datasets.

(4) Evaluating Two Approaches Using Combined Datasets with Added Noise
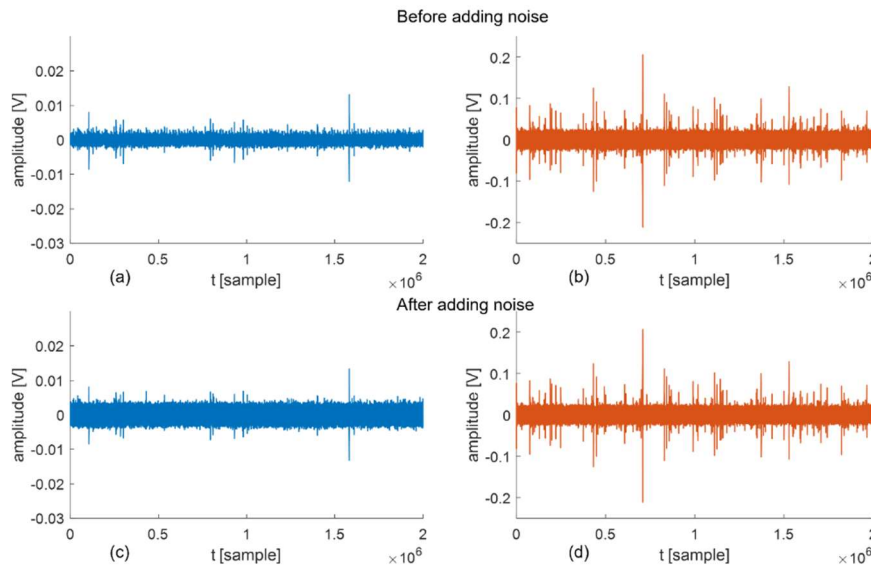
Figure 2. 12. AE on channel 2 before/after adding noise with $\gamma = 0$ dB: (a, c) NORMAL, (b, d) ABNORMAL

To evaluate robustness of the proposed method, Gaussian noise was added to the acquired signals to increase the intensity of background noise. The parameter γ is counted in (dB) for each channel of the three AE sensors. Figure 2.12 illustrates the signals from channel 2 before and after added noise with $\gamma = 0$ (dB). This evaluation was conducted using the k–NN classifier trained by the combination of all datasets. The testing accuracies are shown in Table 2.10.

Table 2.10 shows that if γ is small, the accuracies of the two methods are approximately the same as in Table 2.9. When γ increases, the accuracy of both approaches decreases, and at a

certain value, they cannot guarantee classification. However, the direct AE–based method fails in classification when the added noise reaches 5 (dB) whereas the proposed $g(r)$–based method works until γ reaches 35 (dB). Thus, the $g(r)$–based approach produces a classifier that can detect a small leak more robustly than the direct AE–based classifier.

Table 2. 10. Accuracy with added noise

γ [dB]	Model	
	Direct AE–based	**$g(r)$–based**
-30	99.28	99.97
-20	99.28	99.97
-10	99.28	99.97
-5	99.25	99.96
0	99.08	99.83
5	0	98.46
10	0	94.38
15	0	88.83
20	0	76.9
25	0	70.74
30	0	66.96
35	0	0

2.5. Conclusions

AE–based techniques have the advantages of detecting the physical changes within a structure such as water pipeline with high accuracy. Due to the complication of the AE phenomenon, model–based fault diagnosis of the water pipeline is difficult whereas a data–driven fault diagnosis is relatively easy to implement. However, if the training relies on extracting features directly from AE signals, the resulting classifier is unreliable because such data cannot reflect all possible information under divers working conditions. This paper introduced an intermediate step in a water pipeline fault diagnosis framework in which the signals are preprocessed before the extraction of features. The preprocessing step calculates an intermediate quantity $g(r)$ by applying an attenuation equation to the acquired AE signals. The quantity $g(r)$ is more stable than the original AE signals in leakage identification. The results showed that the $g(r)$–based approach achieves higher accuracy and higher reliability than the direct AE–based method. This approach can be applied in a variety of systems because the $g(r)$ function always possesses the same behavior reflecting leaks in a system. This paper also proposed a way to select a reasonable window size when dividing signal frames. The proposed technique uses the speed of acoustic emission signal to calculate a possible maximum lag time between signals on the sensor channels. From that computation, the frame size can be adjusted appropriately.

Chapter 3

Leak Localization in an Industrial–fluid Pipeline based on Acoustic Emission Burst Monitoring

3.1. Introduction

Since AE waves propagated to sensors along a pipeline comply with the wave dispersion rule of attenuation and velocity [21], AE source localization can be relied on either amplitude (related to the wave attenuation) or arrival time (related to the wave velocity) [47]. The amplitude approach is hard to apply because the attenuation function of AE signals is extremely complicated – they depend on both frequency and operating condition [21, 37, 48]. As a result, conventional methods harnessing arrival time have been used more frequently to pinpoint an AE source in a pipeline [47].

AE source localization based on arrival time estimates the TDOA between AE signals recorded by two sensors. Although the value can be computed by a CCF [30], it is strongly dependent on the chosen wave propagation model and background noise [49]. Moreover, the propagation path is not always straight from the defect to the sensors, and could instead involve reflection, diffraction, or divergence. In addition, AE signals are not from a unique source; they come from multiple sources, such as vibrations at joint points of flange, inner fluent disturbance places, or external random collisions. Waves also propagate in various modes and can superpose at the point of measurement [21]. Consequently, AE source localization via CCF results in considerable error.

A GCC function [50-52] was presented to improve the shape of the CCF, thus obtaining superior accuracy of the TDOA estimator. The CC between two signals is associated with their cross power spectral density function through an inverse Fourier transform and a general frequency weighting function which is a prefilter added into the CCF estimator [52]. The prefilter includes Roth, SCT, PHAT, EF, etc. However, authors in [52] claimed that designing a prefilter needs a priori knowledge of leak signals and noise. Indeed, the issue is challenging in industrial-fluid pipeline systems because countless factors can impact on AE signals, for instance, environmental temperature, inner fluid pressure and flow rate. In consequence, leak localization with GCC does not provide high accuracy.

Some approaches have been introduced to enhance leak localization in fluid pipelines [53, 54]. They used WDD, and EMD to eliminate components which are uncorrelated with leak signal, and then increased an accuracy of the leak location estimation. However, AE signals are

distorted by both the noise presence and the wave attenuation, the presented methods have their own limitation on improving the quality of CCF to retrieve an expected location accuracy because the most intrinsic components decomposed by WDD or EMD are only related to the remains of leak signals attenuated in their propagation path.

As can be seen in Figure 3.1, a recorded signal in an industrial fluid pipeline contains many AE bursts. They can come from abnormal points — leaks, bending-segments, cracks, or collisions in pipelines [3]. The amplitude and arrival time of AE bursts vary with their source's position and strength. Thus, such bursts can provide relevant information about leak location. This paper proposes an ABM method for leak localization. The monitoring mechanism is implemented in several stages. First, AE bursts are detected by adaptive thresholds from individual sensor channel. Second, the bursts are grouped into pairs. If two bursts in a pair come from a source, they provide an appropriate characteristic of wave dispersion related to TDOA and energy. In addition, same-origin bursts reveal a degree of similarity in the frequency domain despite of the attenuation. Hence, another step of ABM is that AE burst pairs are filtered by constraints of their TDOA, energy, and coherence evaluated by a coherence-squared function [55, 56]. A pair of filtered bursts is referred as an event from an AE source. Next, a statistic technique clusters the coordinates of events in a histogram. Finally, the AE burst monitoring program issues a warning alarm through cluster distribution and a priori information about the pipeline system's design. Experimental results show that ABM identifies leaks more precisely than conventional methods using CCF and GCC to estimate TDOA from recorded AE signals or even from their most intrinsic components returned by WDD and EMD.
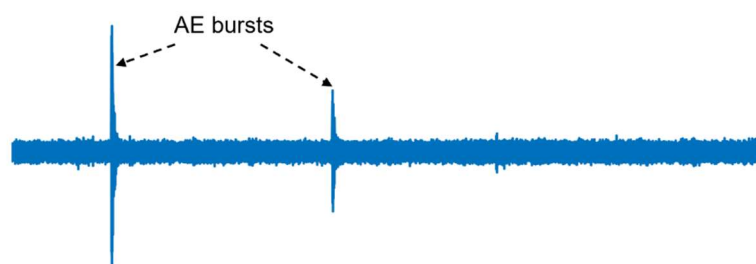
Figure 3. 1. AE bursts.

3.2. Methodology

To illustrate leak localization based on AE burst monitoring, the setup in Figure 3.2 is applied. In Figure 3.2, x–axis is established to coordinate a leak and sensors. As a reference, the origin of x–axis is set at a flange because this position is always permanent. The AE source is referred to as the leak (S) at d [mm] and the AE sensors 1 and 2 are mounted at d_1, d_2 [mm] and designated CH1 and CH2, respectively. Other parameters are also symbolized in the picture.
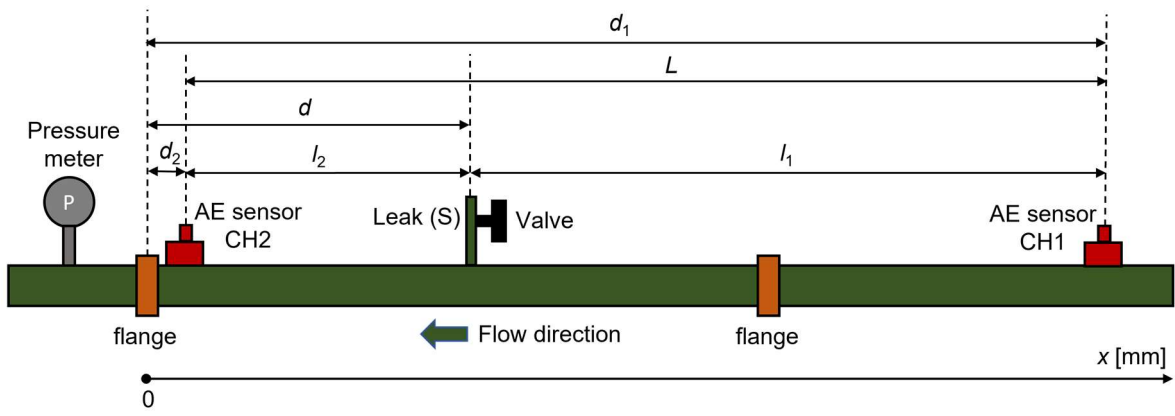
Figure 3. 2. AE source and sensor mounting.

The localization methodology is shown in Figure 3.3.
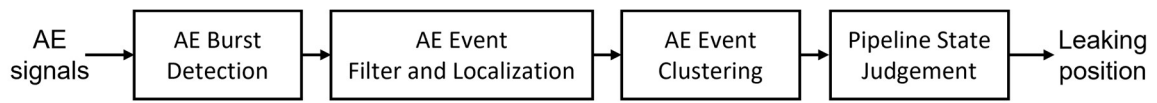
Figure 3. 3. Fault localization based on AE burst monitoring.

3.2.1. AE Burst Detection

Since AE bursts can be detected from the envelope of signal, we start with an AE burst detection process using a Hilbert transform and an envelope detector, as depicted in Figure 3.4. The Hilbert transform is responsible for converting a real input value sequence into an analytic form that is a complex signal. Then, the envelope detector simply takes the modulus of the complex signal at the output of the Hilbert transform block to have the signal envelope. More information about the envelope detection method is available in [56].

Bursts are then searched from the signal envelope on the output of the envelope detector. The Neyman–Pearson theorem [57] states that to maximize P_D (detection probability) for a given $P_{FA} = \alpha$ (false alarm probability), it is necessary to decide if H_1 satisfies

$$L(z) = \frac{p(z|H_1)}{p(z|H_0)} > \gamma \tag{3. 1}$$

where the threshold γ is found from

$$P_{FA} = \int_{\{z:L(z)>\gamma\}} p(z|H_0)dz = \alpha \tag{3. 2}$$

where the function $L(z)$ is the likelihood ratio, H_0 is the null hypothesis (signal absent), H_1 is the alternative hypothesis (signal present), z is an observed set of data, and $p(z)$ is the probability density function. Eq. (3.1) and Eq. (3.2) are the mathematical fundamental of threshold calculation for AE burst detection with a given false alarm probability. The existence of a burst

is the H_1 hypothesis if the envelope value exceeds the threshold, and the case of no burst is the H_0 hypothesis if the signal envelope lies below the threshold. Those expressions are implemented in the threshold calculation block of the AE burst detection process in Figure 3.4. As can be seen in the detection diagram, a threshold is computed from the AE input signal. Since AE signals depend on operating conditions of a pipeline such as fluid pressure, flow rate, and temperature [21], they are constant variation of AE signals. Accordingly, the applied threshold which is continuously updated in time can adapt to the operating condition.
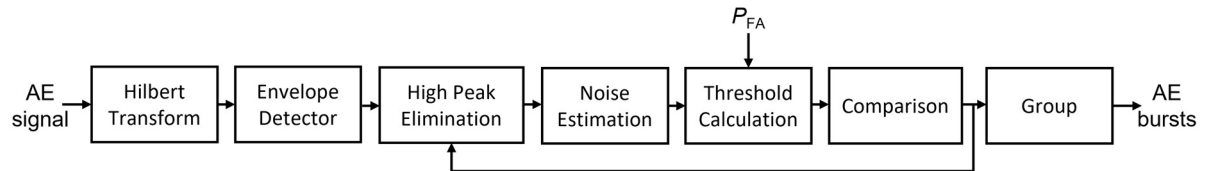
Figure 3. 4. AE burst detection based on adaptive thresholds.
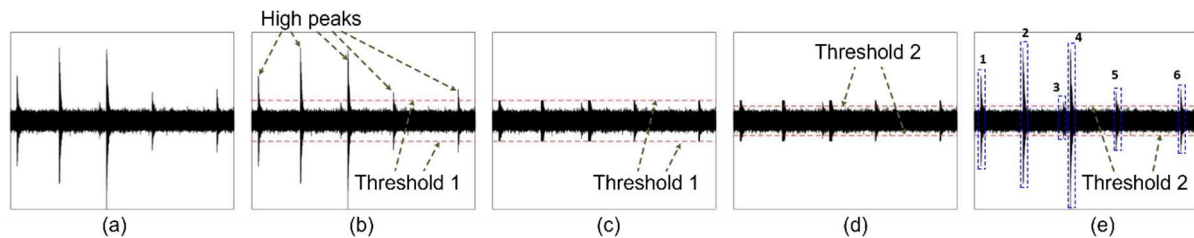
Figure 3. 5. AE bursts detected in a signal: (a) Raw signal, (b) Seeking high peaks, (c) After eliminating high peaks, (d) Calculating threshold 2, (e) Detecting bursts

In the burst detection process, high–peak elimination plays a central role in estimating noise level. If high peaks are involved in the calculation, the estimated noise level differs significantly from the true value. This leads to missing signals and increasing time-of-arrival error. Hence, we use two thresholds in the burst detection, as illustrated in Figure 3.5. We use the first threshold, which is called a coarse threshold, to eliminate high peaks over the noise level in the signal, as shown in Figure 3.5 (c). With the remaining signal after eliminating high peaks, a second threshold (fine threshold) is computed. The second threshold is smaller than the first one; thus, low bursts can be identified, for example of the burst 3 in Figure 3.5 (e), unlike in Figure 3.5 (b). The same samples of the burst 3 cannot be detected because they are completely under the first threshold.

Apart from the adaptive threshold, this paper suggests grouping adjacent bursts as an AE wave is likely possible separated into many bursts due to the diversity of propagation modes. The grouping step aims to unify the bursts, decreasing the false alarm rate for event detection. The constraints of this process are defined by analyzing the wave models and propagation velocities

given in [21, 43]. Figure 3.6 illustrates an AE burst created from neighboring bursts. Here, two detected bursts are considered neighboring if their time distance is smaller than the separability of waves propagated in the lowest and highest velocities from a point in the pipeline. Expression (3.3) establishes the condition for the grouping.

$$t_{neighboring} \leq \left(\frac{l}{C_{min}} - \frac{l}{C_{max}} \right)$$

(3. 3)

Here, $t_{neighboring}$ is the time-distance between the two bursts; l is distance from the burst source to the sensor; and C_{min} and C_{max} are the lowest and highest wave velocities, respectively.
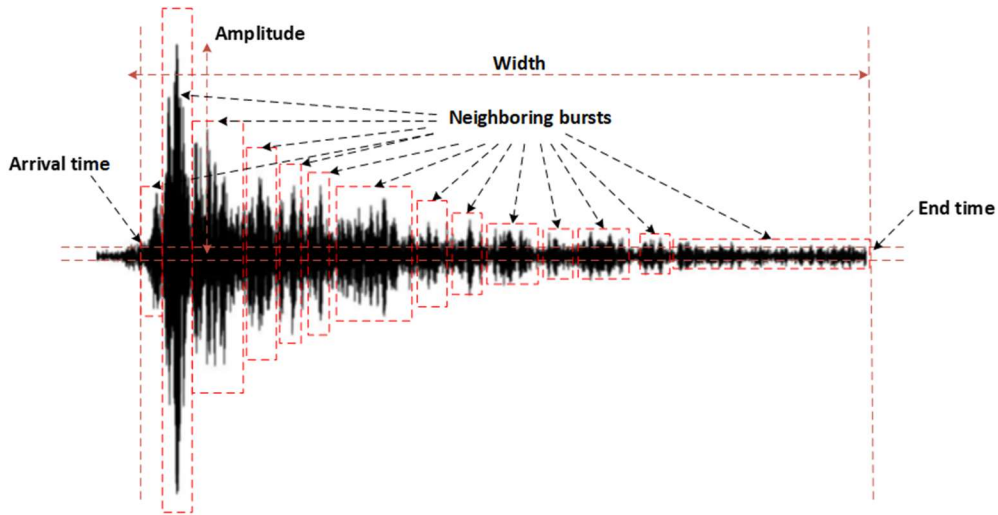
Figure 3. 6. A grouped AE burst formed from neighboring small bursts.

Since AE waves can propagate in the walls of pipes and in the inner fluid [42], C_{min} and C_{max} are considered by comparing speeds in both media. Wave velocity in solids is dominant against fluid velocity [43], and C_{min} and C_{max} values can therefore be considered as the fluid and solid wave velocities, respectively. However, because the source of a burst between CH1 and CH2 may be ambiguous, calculations of neighboring time assume that l is L, which is the distance between the two sensors. This value is the maximum possible path of direct wave propagation from the wave source to one of the sensors in the tested pipeline.

3.2.2. AE Event Filter and Localization

An AE event is defined to be a pair of bursts propagated from the S fault to the CH1 and CH2 sensors as depicted in Figure 3.7. According to the wave propagation theorem [21], the x coordinate is calculated by:

$$x = d_2 + \frac{L - C\Delta t}{2}, \Delta t = t_1 - t_2$$

(3. 4)

where C is the wave propagation velocity and t1 and t 2 are the arrival times. Although expression (3.4) allows for localization of an AE source, they cannot distinguish which pair of bursts is the true event. In a true event, a pair of bursts come from the same source. If they are from two different sources, the pair is useless and should be discarded. This paper offers constraints to filter such events.
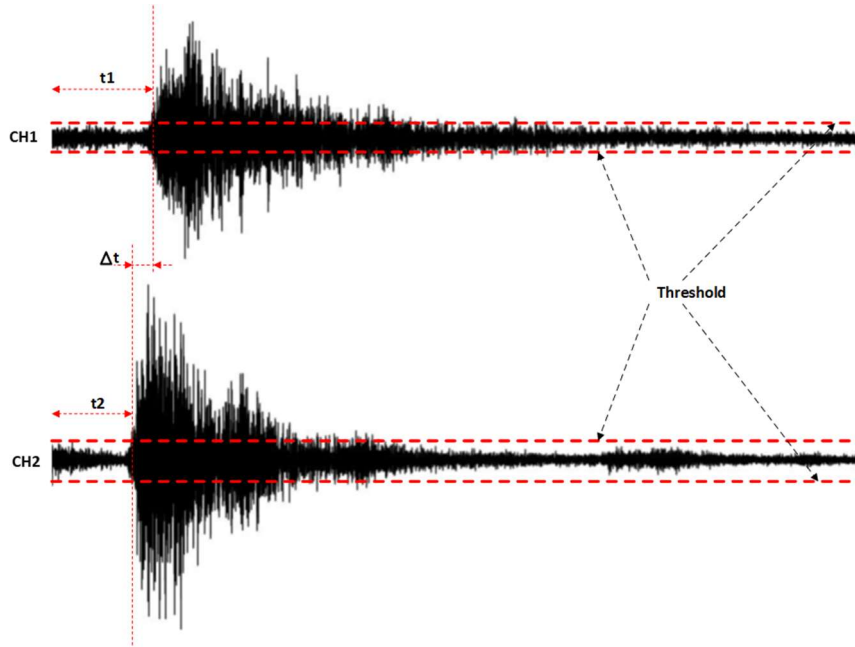
Figure 3. 7. An AE event represented by two bursts in the CH1 and CH2 channels.

First, from the frequency response of the system in [37] we can observe that the energy of an AE wave attenuates in proportion to propagation distance as it disperses along pipeline. A possible pair of bursts can therefore be checked by the attenuation condition:

$$(l_1 - l_2)(E_2 - E_1) \geq 0 \tag{3. 5}$$

in which l_1 and l_2 are the distances from S to CH1 and CH2, respectively, and E_1 and E_2 are the respective burst energies. In Eq. (3.5), if l_1 is greater than l_2, then CH1 receives an amount of signal energy smaller than those of CH2 which receives from the leak signal. Additionally, if two bursts originate from the same source, they must comply with the coherence-squared function [55, 56], which is written as:

$$C_{12}(t,f) = \frac{|S_{12}(t,f)|}{\left[S_{11}(t,f)S_{22}(t,f)\right]^{\frac{1}{2}}} \geq \beta \tag{3. 6}$$

where β is a threshold ($0 \leq \beta < 1$), $C_{12}(t,f)$ is the coherence-squared function estimate, $S_{12}(t,f)$ is the cross-spectral density function estimate, and $S_{11}(t,f)$ and $S_{22}(t,f)$ are spectral density function estimates of the two bursts, respectively. The signals are considered nonstationary

processes, where t is time and f is the frequency. In addition to the impact of pipeline material, fluctuation of operating conditions (inner fluid pressure and flow rate), and surrounding environment, AE signals suffer from inevitable distortion in their dispersion because the wave attenuation is a function of frequency and propagation distance [37, 48]. Therefore, the coherence-squared function can receive a comparatively low value to an AE burst pair even though they are from a source. For example, if leak location is close to CH1 and far from CH2, the coherence-squared function of two bursts acquired by the two sensors provides smaller value compared to the case that the leak is located at the middle of these sensors. However, if the leak location is unknown, and the external factors (pressure, flow rate, and temperature) are constantly variable, we need set β not to be so high to avoid missing useful AE events which come from a leak and not to be so low to eliminate useless AE events including two bursts from two different sources. Thus, an optimal value of β should be chosen in practice. If a large number of satisfactory AE events is collected in a certain time, but their coordinates are randomly scattering which makes difficulty in event monitoring, we can increase β to restrict this phenomenon. In other word, if there are a few events, we can decrease β to get more events.

Furthermore, as the leak is between the two sensors, its coordinate should satisfy:

$$d_2 \leq x \leq d_1 \tag{3. 7}$$

Combining (3.4) and (3.7) produces:

$$|\Delta t| \leq t_{\text{maxtdoa}} = \frac{L}{C} \tag{3.8}$$

Expressions (3.5), (3.6), and (3.8) are constraints for AE event filtering.
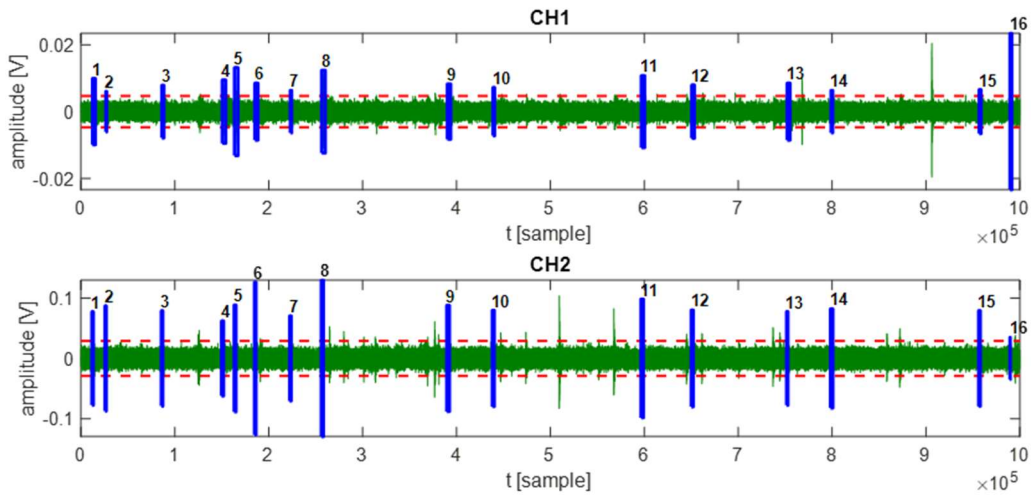
Figure 3. 8. Filtered AE events.

Figure 3.8 includes sixteen filtered AE events marked in blue with indexes above in a frame of one million samples per channel. Some bursts are eliminated even though their amplitudes are not small. Some unmet events can be generated by the leak source; however, they are seriously distorted due to attenuation, interference, or flow disturbance. As a result, they provide no valuable information about the leak location and their absence could enhance the localization.

3.2.3. AE Event Clustering and Judgment

According to the theorem of probability and statistics [56], a decision is made by a dominant probability of event occurence. Hence, the leaking position can be determined at the point where AE events spread more densely than surroundings. This can be seen in Figure 3.9 (a), which depicts the distribution of AE events by coordinates and time. Although the AE events can come from any place on the tested pipeline due to turbulence in the fluid flow, the leaking region would turn out an event distribution denser than its neighboring positions because a leaking location is always non-movable. Therefore, in order to address leak localization, we can rely on the probability of AE event occurence acoording to location along the pipeline. We examine the event occurence probability in regions created by equivalently dividing a pipeline into small segments instead of monitoring in total continuous points on the pipeline. Call n as a number of divided segments, then size of a segment is given by:

$$u = \frac{L}{n} \qquad (3.9)$$

where u is the length of a segment. We asume that the expected location error of leak localization is Δx. This parameter is usually predefined, which is a technical criterion to evaluate the result. Based on the expected location error, the size of segment is selected by:

$$u \leq \frac{\Delta x}{2} \qquad (3.10)$$

Eq. (3.10) can guarantee that the event distribution has enough resolution to distinguish among event distribution densities which are appropriate for the expected location error (see Figure 3.9 (b)). In the real application, since a pipeline is various in length, a relative error is consequently useful to improve the location accuracy of a leak localization algorithm. This error is given by:

$$e_r = \frac{\Delta x}{L} \text{x}100 \qquad (3.11)$$

where e_r refers to the relative error in percentage. From Eqs. (3.9), (3.10), (3.11), we can determine the number of segment by (3. 12).

$$n \geq \frac{200}{e_r} \qquad (3.\ 12)$$

Eq. (3.12) is a mathematical expression to choose the number of segments to quantify the event distribution density. Due to the lack of a priori information about a probability density function of leaking events, we assume that they have a uniform distribution. It means that the probability density is the same for all the events existing in segments. Thus, the probability of event occurence in a segment is proportional to event distribution density, which is equal to sum of events in the segment. The leak location is therefore determined through density peaks in the event distribution density, as illustrated in Figure 3.9 (right), where the events are arranged in proper segments depending on their coordinates matched in the *x*-axis along the pipeline. In this graph, the leak is located at the point of the highest density.
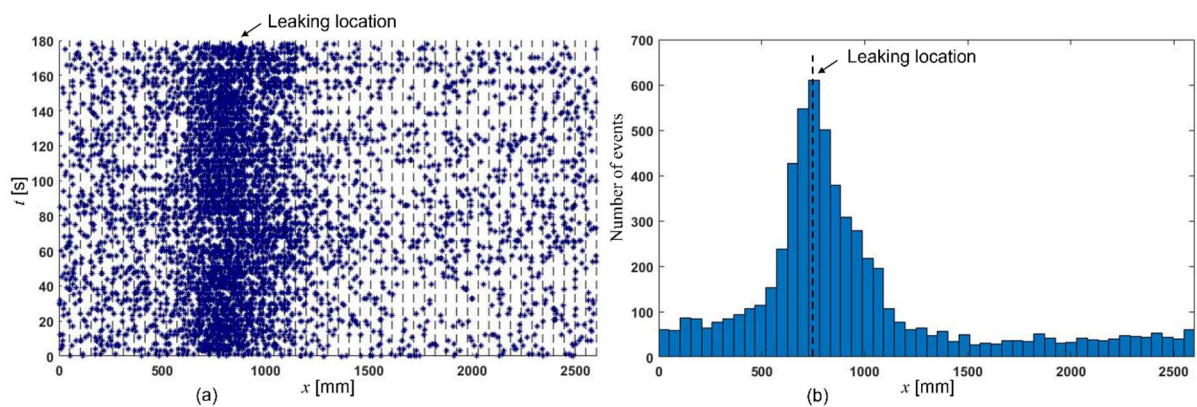
Figure 3. 9. AE event clustering: (a) Event occurrence over time and coordinate, (b) Source distribution density

3.3. Experiment Setup

3.3.1. Testbed Configuration
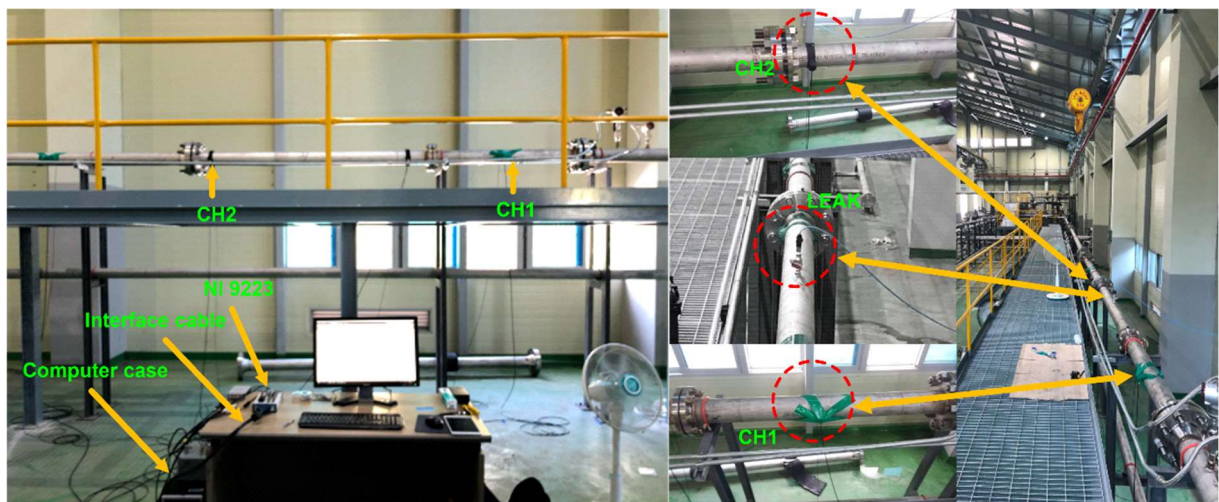
Figure 3. 10. The pipeline testbed.

Figure 3.10 shows a testbed and an AE equipment system to record AE signals in a tested part of a water pipeline system. The installation is matched with Figure 3.2 and the experimental parameters are listed in Table 3.1.

Table 3. 1. Experimental parameters

No	Quantity	Detail
1	Location of sensor 1 (d_1)	2600 [mm]
2	Location of sensor 2 (d_2)	100 [mm]
4	Location of leak (d)	900 [mm]
5	Thickness of pipelines	6.02 [mm]
6	Outer diameter of pipelines	114.3 [mm]
7	Material of pipelines	Stainless steel 304

The leak was designed by a tool welded into a hole on the pipe wall. Four leaks with diameters (2.0, 1.0, 0.5, 0.3) [mm] are abbreviated by $F = (F_1, F_2, F_3, F_4)$, respectively. There is a valve on this tool to activate or deactivate the leakage simulation in the system.

3.3.2. Acquisition Equipment System
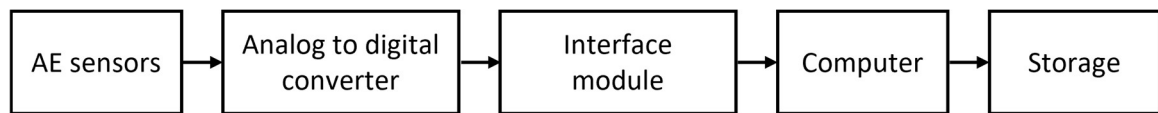
Figure 3. 11. The overall paradigm of AE equipment system.

Figure 3.11 illustrates an overall process of the AE equipment system for data acquisition in this study. R15i–AST sensors from MITRAS corporation were used for data acquisition because they provide high sensitivity. The 16–bit analog to digital converter with controllable sampling frequency and the interface module through the high–speed universal serial bus standard are integrated in a NI–9223 module manufactured by NI. The computer is a personal computer compatible with NI–9223 interface, and a hard drive of one terabyte is installed as a data storage. The AE sensors were attached to the surface of a pipeline using a set of mounting kit and tapes to fix their positions. In order to reinforce the contact between the sensor and the pipe surface, we added a type of specialized gel into the contacting area. After finishing the setup of hardware, a data recording software which is self–developed using a Python language and an interface library supported by NI was used to control the data acquisition system. Then, the AE equipment was tested by the pencil lead break method [58]. This ensures a suitable sensitivity for the data acquisition system and calibrates the sensors, thus resulting in reliable AE signal datasets. The operating principle of the system in Figure 3.11 is composed of several phases. First, AE waves are converted to electronic signals by an AE sensor. Since the electronic signal in an analog form, it is converted into digital signal by the ADC. Second, the interface

module is responsible for communicating between the ADC and the computer. Finally, the data recording program receives data from the interface module and saves them in a hard drive.

3.3.3. Data Record
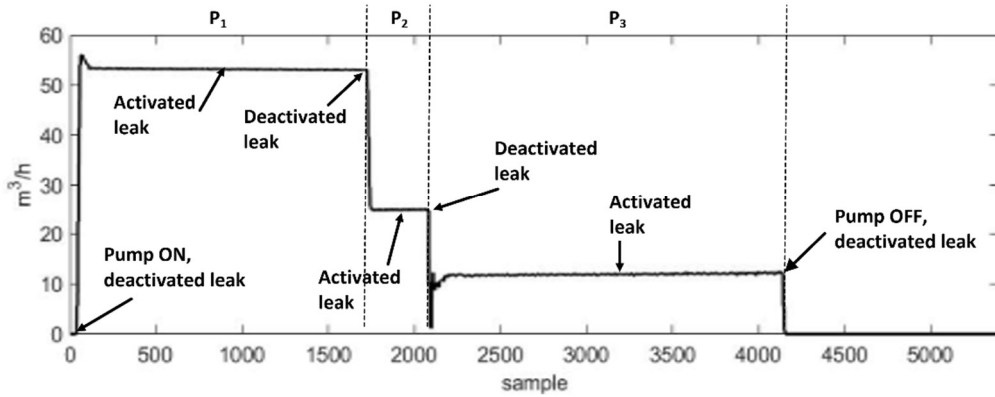
Figure 3. 12. The water flow rate in the F_1–leak experiment.

Datasets were collected in the leaking state, referred to the open position of the valve, the water flow was controlled in pressures (7, 13, 18) bar called $P = (P_1, P_2, P_3)$, and the environmental temperature was roughly 26° Celsius degree. In a conditional pair F_iP_j ($i = 1, 2, 3, 4$; $j = 1, 2, 3$), a dataset was recorded for two minutes with a sampling rate of 1 MHz. Figure 3.12 presents the experimental process with the F_1–leak. In the beginning, the pump was turned on, the valve was closed (deactivated leak), and the pressure was adjusted to P_1. The valve was then opened (leak activated), with the recording device waiting for flow stability before acquiring data for two minutes for the pair F_1P_1. Next, the experiment continues with the pressures P_2, P_3, as depicted in Figure 3.12. Prior to switch conditions, the leak is deactivated to standardize the moment of pressure measurement when the pipeline is normal. After completing the F_1–leak experiment, the data acquisition with the leaks F_2, F_3, F_4 were repeated.

3.4. Experimental Results

To evaluate the proposed method, the approaches using CCF along with WDD, EMD and GCC with PHAT were implemented on the same datasets to which ABM was applied. WDD and EMD techniques were employed to search for the most intrinsic components from the raw AE signals. After decomposing the original signals into distinct components for both channels, they were grouped in corresponding pairs to calculate the cross-correlation. The one that returned a maximum correlation coefficient was assumed to estimate TDOA.

The window size was given below, according to the suggestion in Subsection 2.3.3.

$$t_{\text{frame}} = 10\Delta t_{lag}, \quad \Delta t_{lag} = \frac{L}{C} \tag{3.13}$$

In (3. 13), t_{frame} is signal frame size, Δt_{lag} is the maximum lag time, and C is the wave velocity in water. This equates the lag part of signal to 10% of the whole frame.

Table 3.2 lists the necessary parameters used in this study, which were calculated from a priori information about the experimental setup. As we digitized AE signals, the values t_{frame}, t_{maxtdoa}, and $t_{\text{neighboring}}$ with unit of second can be represented by N_{frame}, N_{maxtdoa}, and $N_{\text{neighboring}}$ with unit of sample. The parameters α and β were manually selected to get the expected effectiveness of event filtering, hence reducing the relative location error.

Table 3. 2. Implementation parameters

No	Parameter	Value
1	Frame size (N_{frame})	16700 [sample]
2	Wave velocity (C)	1,500,000 [mm/s]
3	Expected relative error (e_{r})	4%
4	Number of divided segments (n)	50
5	The maximum TDOA (N_{maxtdoa})	1667 [sample]
6	False alarm probability (α)	0.1%
7	The coherence threshold (β)	0.8
8	Time–distance ($N_{\text{neighboring}}$)	1667 [sample]

Multiple peaks appeared in the coordinate distribution density, as depicted in Figures 3.13.c, 3.14.c, 3.15.c, and 3.16.c. This can be attributed to the fact that the correlators implemented by CCF or GCC were continuously applied to all the signal frames while intrinsic components of the leak could not be correctly extracted, resulting in the low correlation coefficients of Figures 3.13.b and 3.16.b. Although Figures 3.14.b and 3.15.b present higher coefficients due to the application of WDD and EMD, these values are still low and leak localization was still ineffective. However, the AE sources in Figure 3.17 largely concentrate around the leak position for multiple datasets in diverse conditions as the localization has been deliberately applied only to the filtered events. With the AE event filtering technique as presented in Subsection 3.2.2, unwanted events are excluded, improving the accuracy of leak localization.

Table 3.3 shows the leak coordinates and their relative errors resulting from all the datasets if the leak position is selected to be the place with the highest density. The average errors in this table claim that the ABM method is more effective than the others. Although CCF or CCF with EMD return the nearly correct location of the leak in some datasets (for example, F_2P_1) and the errors are not greater than 2%, similar results were not provided in other cases, such as F_4P_1.

This can be explained by the low correlation coefficients, as illustrated in Figures 3.13.b, 3.14.b, 3.15.b, and 3.16.b. In contrast, the ABM method identifies the leak signals through obvious signatures before the localization, which was presented in Subsection 3.2. This approach can therefore enhance the results.
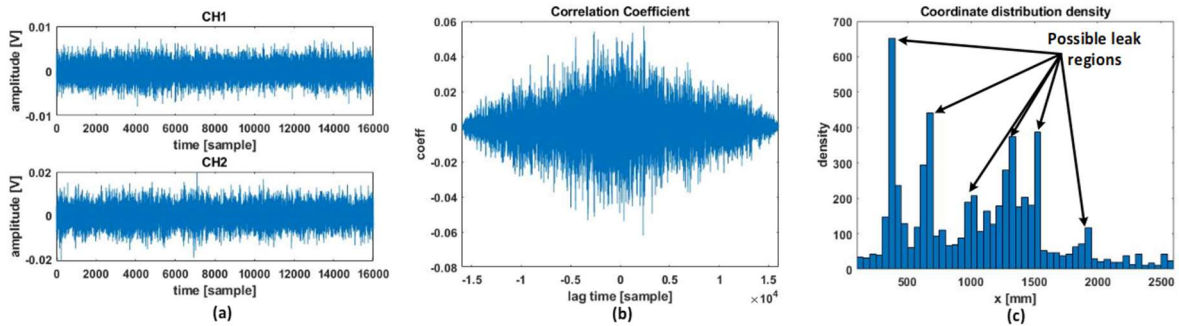
Figure 3. 13. Leak localization using CCF for the F_1P_2 dataset: (a) raw signals, (b) CCF, (c) density.
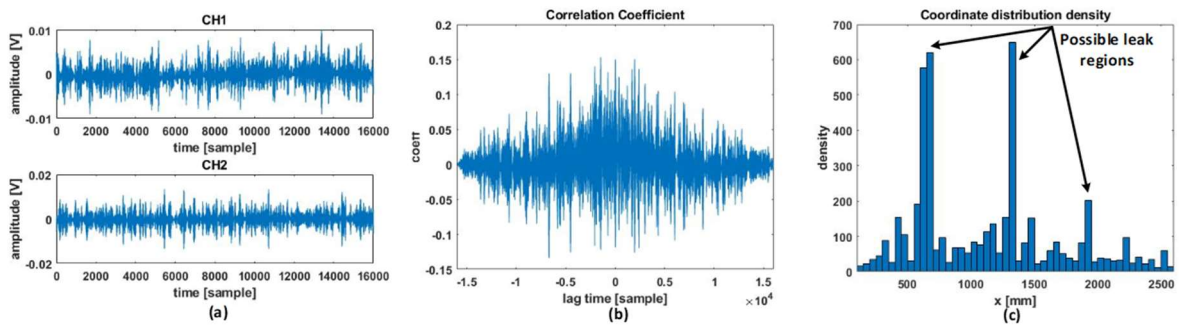
Figure 3. 14. Leak localization using CCF combined with WDD for the F1P2 dataset: (a) raw signals, (b) CCF, (c) density.
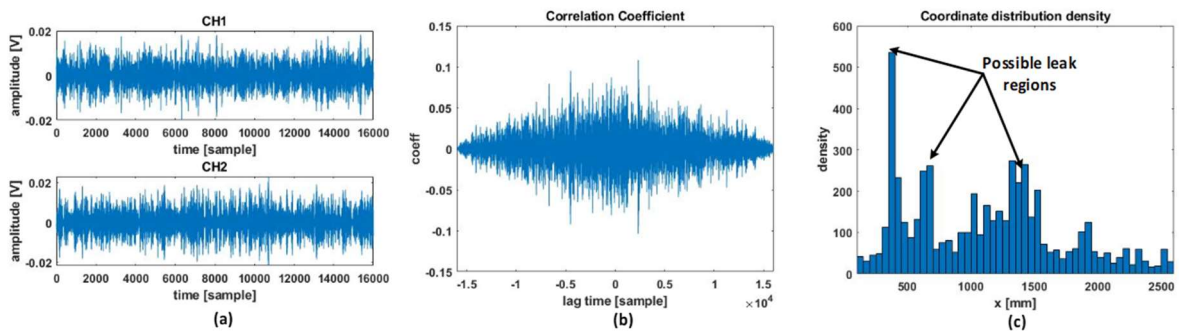
Figure 3. 15. Leak localization using CCF combined with EMD for the F1P2 dataset: (a) raw signals, (b) CCF, (c) density.
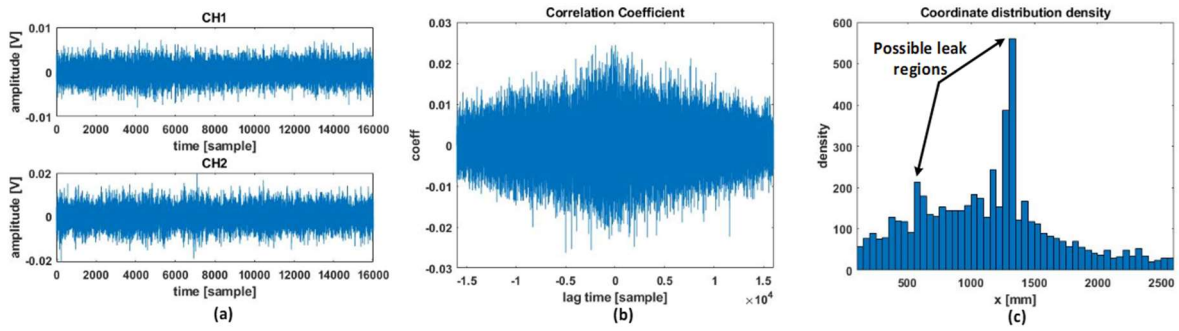
Figure 3. 16. Leak localization using GCC through a PHAT function for the F1P2 dataset:(a) raw signals, (b) CCF, (c) density.
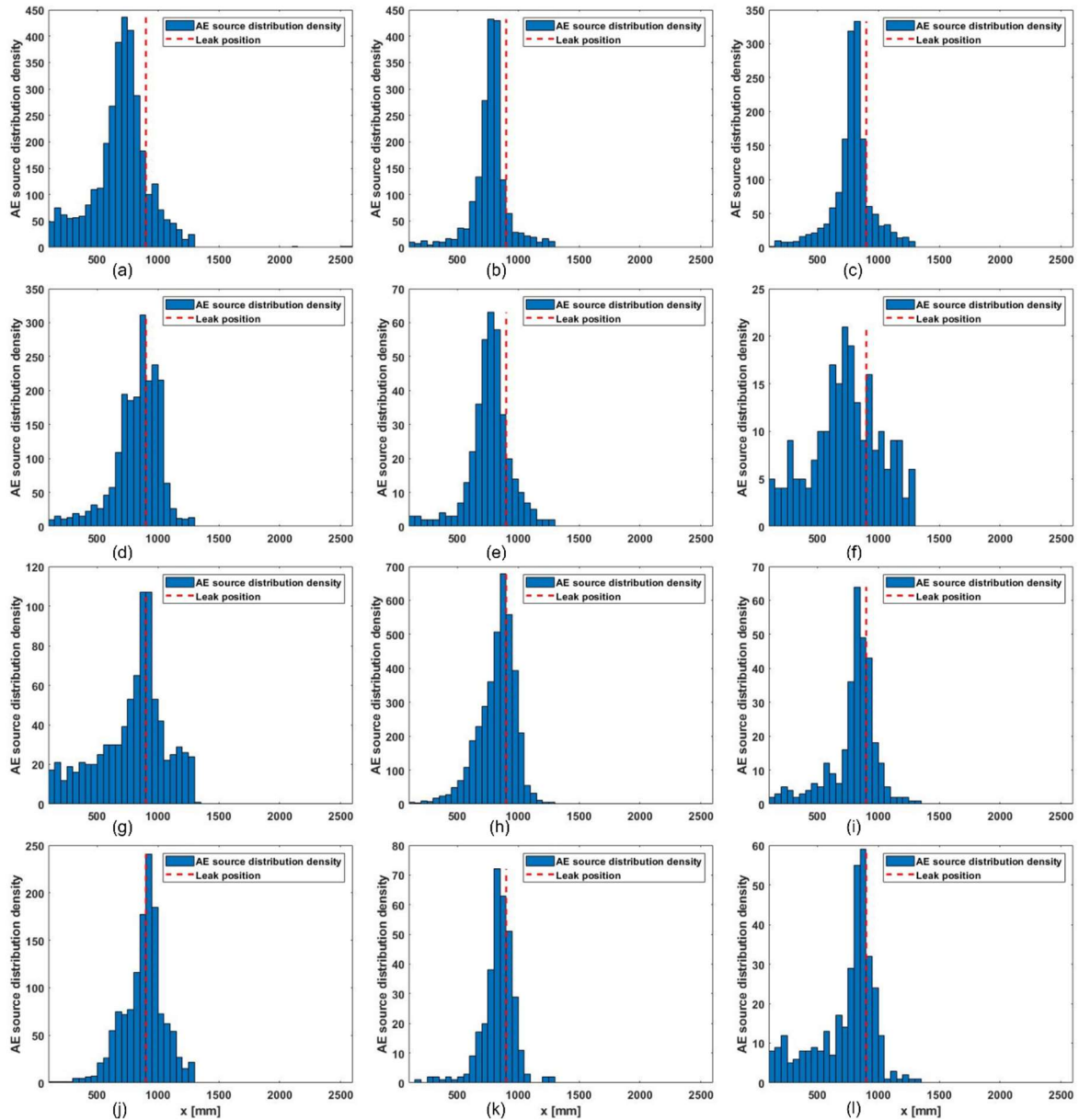
Figure 3. 17. AE source distribution density using ABM: (a) F_1P_1, (b) F_1P_2, (c) F_1P_3, (d) F_2P_1, (e) F_2P_2, (f) F_2P_3, (g) F_3P_1, (h) F_3P_2, (i) F_3P_3, (j) F_4P_1, (k) F_4P_2, (l) F_4P_3

Table 3. 3. Leak coordinate *x* [mm], relative error [%]

		CCF	CCF with WDD	CCF with EMD	GCC with PHAT	ABM
				Method		
Dataset	F_1P_1	1325; 17	1325; 17	1275; 15	1325; 17	800; 4
	F_2P_1	950; 2	1025; 5	875; 1	1100; 8	850; 2
	F_3P_1	800; 4	1325; 17	1325; 17	1325; 17	900; 0
	F_4P_1	1400; 20	1325; 17	1325; 17	1500; 24	950; 2
	F_1P_2	400; 20	1325; 17	1325; 17	1325; 17	800; 4
	F_2P_2	850; 2	1325; 17	975; 3	1325; 17	800; 4
	F_3P_2	1100; 8	1325; 17	875; 1	1325; 17	850; 2
	F_4P_2	1325; 17	1325; 17	1325; 17	1325; 17	800; 4
	F_1P_3	1350: 18	1325; 17	1425; 21	1325; 17	850; 2
	F_2P_3	1400; 20	1125; 9	625; 11	1325; 17	800; 4
	F_3P_3	450; 18	1325; 17	1075; 7	1325; 17	850; 2
	F_4P_3	1250; 14	1325; 17	1375; 19	1325; 17	900; 0
Average		1022.7; 13.3	1283.3; 15.3	1150; 12.2	1320.8; 16.8	845.8; 2.5

3.5. Conclusions

Leak localization using the proposed acoustic emission burst monitoring outperformed conventional methods. The methodology was applied to an industry in which pipelines are typically made of steel to handle adverse conditions. In this environment, acoustic emission signals can be distorted by complications in wave propagation and are prone to interference from ambient noise. Conventional techniques such as cross-correlation function and generalized cross-correlation — even when combined with Wavelet decomposition and denoising and empirical mode decomposition — cannot produce sufficiently accurate location information. However, the experimental results showed that leak regions exhibit the highest density of acoustic emission sources using the proposed method, facilitating leak localization in industrial-fluid pipelines. This is because the proposed method employed a burst detection technique using adaptive thresholds in which they are automatically updated depending on measured acoustic emission signals. This burst detection method offers peak elimination to boost background–noise estimates which converge to the true value and enhance detection precision. Moreover, a grouping tactic helps reduce the false alarm rate associated with wave separation. The event filtering revealed the coordinate concentration of leak events in a clustering histogram, further improving localization accuracy.

Chapter 4

Real–time Leak Detection for a Gas Pipeline using Hybrid Acoustic Emission Features and a *k*–Nearest Neighbors Classifier

4.1. Introduction

Researchers tend to adopt a data–driven approach that trains a classifier using AE features extracted from AE signals to separate pipeline health states to normal or leaking. This approach is appropriate because an AE signal acquired from a gas pipeline is non–stationary [26, 59]. Moreover, AE waves attenuate along the pipeline from their emission source to AE sensors[37]; they vary with the environmental conditions of pressure, flow rate, and temperature [21]. Thus, it is challenging to draw an explicit model to identify a leak relying on AE signals exclusively. A classification model learns the leakage manifestation from the supplied training data; hence, it can identify the leak detection problem effectively. However, the computational complexity of existing leak detection methodologies restricts their exploitation in real–time applications, despite the fact that they show high classification accuracy. For example, the Wavelet transform and the signal decomposition algorithms are used to analyze AE signals, and machine learning-based models are used for state classification [17, 19, 22, 23], which can improve accuracy, but their computation is highly complex.

A long gas transportation system usually comprises numerous pipeline segments with diversity in size, shape, and material. Many sensors are spread over that system to monitor the health of different pipeline segments. A wireless-based leak detection system with a server receiving and exploring signals dispatched from remote sensor nodes, as proposed by [60], would not be suitable for AE signal application due to the overload of communication and computation. Therefore, a sensor node should be a smart integrated system that can itself inspect a pipeline segment and report only the health state of pipeline to its server instead of sending a massive amount of AE signals to the server. The advantage of the integration is that it does not require a complex communication network topology between the sensor nodes and the server. Nonetheless, the integrated system must be low-power and compact, because if many devices are installed, they will result in high energy consumption and a bulky system. This is similar to the design presented in [61], which integrated a propane sensor with a low-power system-on-chip device. However, a propane sensor could only detect an obvious gas leak nearby, thus challenging the early gas leakage detection in a large pipeline network, where a tiny leak would occur at any place and any time.

Working from the demand for gas pipeline leak detection and the achievements and limitations of current studies, this work presents an MCU–based system designed to diagnose leakage for a gas pipeline in real–time. The system analyzes AE signals locally to identify a leak and just issues a warning of state changing. Because an MCU–based system only supports a restricted resource in memory and execution speed for computing implementation, the paper exploits a k–NN classifier trained by using hybrid AE features directly extracted from raw AE signals. The k–NN algorithm can execute on a limited-resource platform in real–time because it is made up of simple computations and neighbor-searching loops. To optimize the algorithm further, a filtering technique is exploited to remove the least useful elements from the feature pool relying on the three-sigma rule [62] and the KL distance [40], which reduces the number of computation cycles and loops in the correspondingly implemented program, thus accelerating the proposed detection system. The selected features are normalized as well; hence, a trained k–NN model can be applied to various sensor nodes along a pipeline network. Moreover, the trained model can be updated in run–time to adapt to a sensor installation location or any change in the working conditions.

Prior to implementing the leak detection program on an MCU–based hardware platform, the proposed methodology is offline synthesized using the Matlab 2019a software and AE signal datasets recorded at a gas pipeline testbed under diverse experimental scenarios. Thus, the essential parameters of the k–NN classifier (training features and number of nearest neighbors) are chosen to ensure not only the real–time characteristic, but also high accuracy of the leak detection program. Aside from ambient noise, any external factor that can cause the vibrations in the pipeline can trip AE signals. For instance, a random pipe collision triggers a mechanical vibration that generates plentiful elastic waves propagating through the pipeline. AE sensors with enough sensitivity can capture signals resulting from those elastic waves, thus interfering with measured target signals. Hence, a k–NN classifier based on AE signals is subjected to discrete events near the testing pipeline, generating false alarms. To address this problem, the current work proposes monitoring the ALEOR from the output of the state classifier. A final decision of pipeline health state is based on the comparison between the instant ALEOR and a defined threshold, hence avoiding a false alarm.

Finally, the work evaluates the gas pipeline leak detection system constructed from the proposed methodology on the 32F746G–DISCOVERY board (STMicroelectronics, Quakertown, Pennsylvania, USA) using recorded AE signal datasets. Experimental results demonstrate that the system can identify a leak in real–time with high average classification

accuracy under various pressure conditions, and its robustness is satisfactory, even with adding white noise to the input AE signal. Hence, the proposed MCU–based system is applicable for gas leak detection in real applications.

4.2. AE Signal Data Acquisition

A pipeline testbed is established to simulate the gas leakage as shown in Figure 4.1. The testbed is a part of a real gas pipeline system (see Figure 4.1 (c)) made from stainless steel 304 pipelines with sizes of 114.3 millimeters (mm) and 6.02 mm in outer diameter and wall thickness, respectively. To create various leaks, we designed a leak tool as shown in Figure 4.1 (a), which is assembled to the testing pipeline. This tool is composed of a valve and an orifice of diameter 0.3 mm, 0.5 mm, or 1 mm (see Figure 4.1 (b)). Hence, the normal/leaking states of the pipeline are connected to closed/open valve positions.
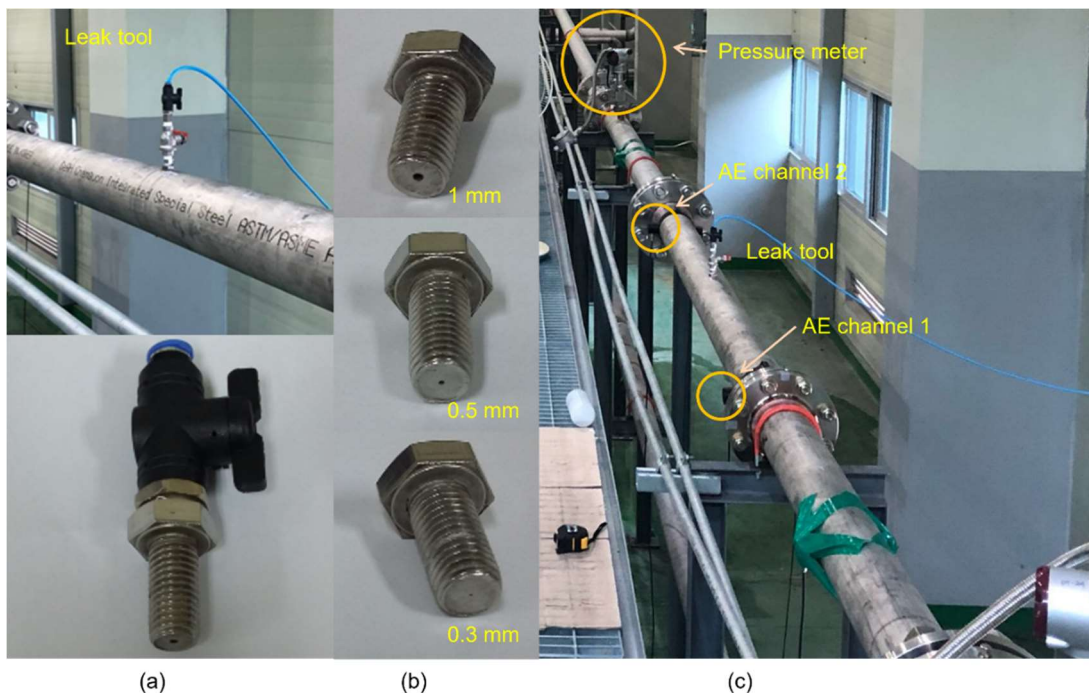
Figure 4. 1. Pipeline testbed: (a) leak tool, (b) orifices, (c) test section.

The experimental configuration is shown in Figure 4.2. To capture AE signals, two R15i–AST sensors (AE channels), which were manufactured by MITRAS Group, Inc (Princeton Junction, NJ, USA), are mounted at downstream and upstream locations on the surface of the testing pipeline. These sensors can detect any elastic wave in a range of operating frequencies, which are 50 kilohertz (kHz) to 400 kHz [41]. Those elastic waves can be caused by diverse sources such as leak noise [17], negative pressure wave [5], ambient noise, and other vibrations of the pipe wall. Such R15i–AST sensors are selected because their operating frequency range covers the frequency ranges of AE waves propagating in metal objects, which are from 100 kHz to 300

kHz, as stated in the BSI standard BS EN 15856 [42]. AE signals are sampled at 1 megahertz (MHz) by the NI 9223 module. The sampling frequency of 1 MHz is more than double the maximum operating frequency of sensors, thus satisfying the Nyquist–Shannon sampling theorem [63] about converting analog signals into digital signals.
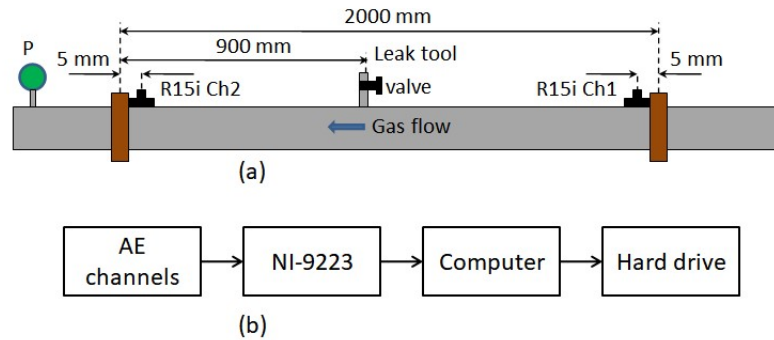
Figure 4. 2. Experimental setup: (a) test section, (b) data acquisition system. (R15i Ch1 and R15i Ch2 are acoustic emission (AE) channels, P is a pressure meter).

After finishing the hardware setup, data recording software is installed on the computer to control the whole data acquisition. Additionally, we exploit the pencil lead break technique [58] to examine both sensitivity of sensors and the whole AE equipment. This ensures the reliability of AE signal datasets prior to storing them in the hard drive.
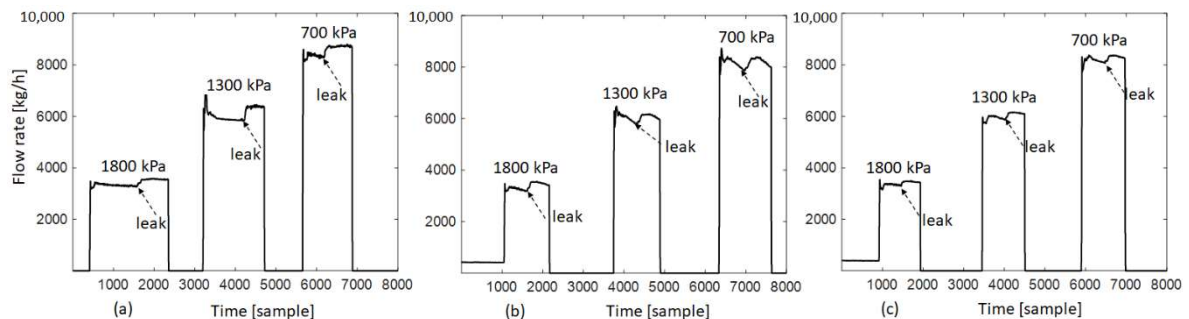
Figure 4. 3. Gas flow rates corresponding to three orifices: (a) 0.3 mm, (b) 0.5 mm, (c) 1.0 mm.

In the experiment, the three orifices are alternated to simulate different leakages at three inner relative pressures of 700 kPa, 1300 kPa, and 1800 kPa, resulting in three normal states of the testing pipeline (closed valve) and nine diverse leaking states (open valve). Specifically, data acquisition has been performed as follows. First, an orifice was installed, and the pipeline system was configured at a pressure level of 700 kPa, 1300 kPa, or 1800 kPa, and this condition was kept relatively stable before acquiring AE signals. At this time, the valve of the leak tool was closed to simulate the normal state of the pipeline. For this state, the signals were recorded for 2 mins. Next, the valve was opened to simulate a leakage. Here, the data corresponding to a leaking state were collected after pressure stabilization. Figure 4.3 presents gas flow rates measured in front of the testing pipeline during the experimental stages.

4.3. Leak Detection Methodology

The overall gas pipeline leak detection diagram is shown in Figure 4.4. It is composed of two processes: one is offline, and the other is online. The offline analysis synthesizes and optimizes the leak detection algorithm, while the online process experiments and verifies the detection. We will describe the analysis blocks of the algorithm below.
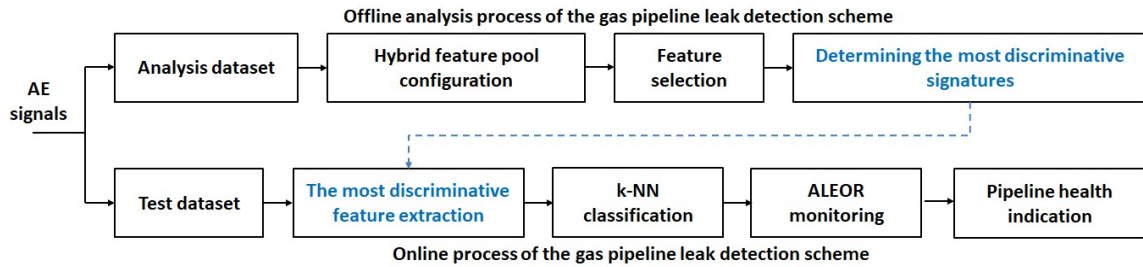
Figure 4. 4. Entire flow diagram of the gas pipeline leak detection.

4.3.1. Hybrid Feature Pool and Feature Selection

To detect the leaking state of a gas pipeline, time and frequency domain statistical features are extracted, as shown in Table 4.1, from raw AE signals utilized as diagnosis leakage signatures. We therefore obtain a hybrid feature pool of size $R \times M$, where R is the number of feature types ($R = 12$, as shown in Table 4.1), and M is the number of analyzed signal frames. The value M should be large enough to reflect the statistical discrimination of the pipeline states precisely.

Next, the feature pool should be refined to enhance the pipeline health classification quality. Outliers, data points that differ significantly from the other aggregated data points in the same class can cause serious problems in statistical analyses. The existence of outliers in a feature extracted from an AE signal measured at a gas pipeline is inevitable, resulting from both exterior and interior factors. The exterior factor could be variability in the measurement. For example, power spikes can interfere with sensed signals, causing outliers in AE features. This problem can be fixed by perfect experimental configuration and the exploitation of high-quality equipment. Outliers may be created by interior factors of the pipeline system, such as burst emissions appearing in high amplitude and energy in AE signals. A gas pipeline itself generates such a signal due to the disturbance between inner gas flow and the gas flow–pipe wall interaction. Nevertheless, outliers should be eliminated from features used for training a classifier because they do not statistically characterize the normal/leaking state discrimination, thus leading to the deterioration of the classification performance. This paper assumes a normal distribution for the AE features; outliers can therefore be detected by the three–sigma rule [62]; this rule is expressed in (4. 1).

$$\Pr\left(\left|Y_i - \mu_{y_i}\right| \leq 3\sigma_{y_i}\right) \approx 0.99 \tag{4.1}$$

where Y_i is an observation from a normally distributed feature y_i; μ_{yi} and σ_{yi} are the mean and standard deviation of the distribution, respectively; $i = 1, 2, \ldots, R$. According to (4.1), if $|Y_i - \mu_{yi}| > 3\sigma_{yi}$, the value Y_i is considered an outlier and it is removed from the set of y_i-feature observations. After unwanted observations are eliminated from the y_i-vector, the length of y_i-vector is shrunk as M_i^* ($M_i^* \leq M$). Because the feature types distribute dissimilarly, the outlier elimination might return different lengths M_i^* of the y_i-vectors ($i = 1, 2, \ldots, R$). As a result, we compensate new satisfactory observations for the feature pool to gain $M_i^* = M$. The feature pool size is therefore intact ($R \times M$); however, its elements are refined, which satisfies (4.1).

Table 4. 1. Typical features for leak detection.

Feature	Equation	Feature	Equation	Feature	Equation
STE	$\sum_{n=0}^{N-1} x_n^2$	STD (σ)	$\sqrt{\dfrac{1}{N-1}\sum_{n=0}^{N-1}(x_n - \mu)^2}$	SKE	$\dfrac{1}{N}\sum_{n=0}^{N-1}\left(\dfrac{x_n - \mu}{\sigma}\right)^3$
RMS	$\sqrt{\dfrac{1}{N}\sum_{n=0}^{N-1} x_n^2}$	ZCR	$\dfrac{1}{N}\sum_{n=1}^{N-1}\left|\text{sign}(x_n - \mu) - \text{sign}(x_{n-1} - \mu)\right|$	SPP	$\arg\max_{f}\left(X(f)\right)$
AVA	$\dfrac{1}{N}\sum_{n=0}^{N-1}\left|x_n\right|$	ETY	$-\sum_{n=0}^{N-1} q_n \log_2 q_n$	SPC (f_c)	$\dfrac{\sum_{m=0}^{M-1} f_m X_m^2}{\sum_{m=0}^{M-1} X_m^2}$
MEA (μ)	$\dfrac{1}{N}\sum_{n=0}^{N-1} x_n$	KUS	$\dfrac{1}{N}\sum_{n=0}^{N-1}\left(\dfrac{x_n - \mu}{\sigma}\right)^4$	SPS	$\sqrt{\dfrac{\sum_{m=0}^{M-1}(f_m - f_c)^2 X_m^2}{\sum_{m=0}^{M-1} X_m^2}}$

where x is an input signal, N is the total number of samples, X is the short-time spectral amplitude, f is the frequency, M is the total number of discrete frequencies, and $q_n = x_n^2 / \sum_{n=0}^{N-1} x_n^2$.

Furthermore, all the extracted features may not be equally effective in highly accurate leak detection. Inferior signatures not only impair the classification accuracy but also increase the computational complexity. Thus, we need to filter out redundant features from the pool to enhance the detection performance while reducing the computational load. This paper scores features using the KL distance [40] and eradicates low–ranked elements in the feature pool. The KL distance is calculated given by (4. 2).

$$d_{KL} = D_{12} + D_{21}; D_{12} = \sum p\left(y_i \middle| w_1\right) \ln \frac{p\left(y_i \middle| w_1\right)}{p\left(y_i \middle| w_2\right)}; D_{21} = \sum p\left(y_i \middle| w_2\right) \ln \frac{p\left(y_i \middle| w_2\right)}{p\left(y_i \middle| w_1\right)} \tag{4.2}$$

where d_{KL} is the KL distance, w_1, w_2 are two classes indicating the normal and leaking states,

respectively; $y_i = [y_{i1}, y_{i2}, \ldots, y_{iM}]^T$ is a sort of y_i-feature in the refined feature pool, p is a conditional probability density function. Based on (4.2), we retain features with the dominant KL distance and remove the others in the feature pool, because the greater the KL distance is, the more discriminative the feature. Finally, we retrieve a purified feature pool with size $r \times M$, where r is the number of high-scored features ($r \leq R$).

4.3.2. Leak Detection Using a k–NN Classifier and Accumulative Leaking Event Occurrence Rate

With the purified feature pool, we utilize a k–NN classifier to distinguish the two normal/leaking states, in which an obscure new class is assigned to the most common class among its k nearest neighbors using the Manhattan distance given by:

$$\delta_j = \sum_{n=1}^{r} \left| z_n - y_{n,j} \right| \tag{4.3}$$

where δ_j is the Manhattan distance between the input feature vector $z = [z_1, z_2, \ldots, z_r]$ and the jth training feature vector $y_{*j} = [y_{1j}, y_{2j}, \ldots, y_{rj}]$, and $j = 1, 2, \ldots, M$. The k–NN classifier categorizes the input z into the major class in its k nearest neighbors corresponding to k minimum distances δ_j ($k < M$).

The detection approach aims at the extremely noisy industrial environment. A k–NN classifier is sensitive to noise involving ambient noise and discrete events and may subsequently yield a false alarm (the classified state is "leaking" but the true state is "normal") or miss a true leaking event (the leakage is actually happening); thus, a normal/leaking state decision should depend on monitoring the ALEOR. The leak detection criterion is given by:

$$ALEOR = \frac{\Delta B}{\Delta t} \geq \gamma, \quad \Delta t = t_2 - t_1 \tag{4.4}$$

where ΔB is the number of leaking events in a time period $\Delta t = t_2 - t_1$, which is from the moment t_1 to the moment t_2, and γ is a threshold to issue a warning of pipeline health state. This threshold is flexibly adjusted by pipeline operators in their specific real environment.

4.4. Implementation of Proposed Gas Pipeline Leak Detection on an MCU-Based Architecture

4.4.1. Offline Analysis of AE Signal Datasets

Prior to developing the real–time gas leak detection program with the proposed methodology

on an MCU–based architecture, we analyzed offline AE signal datasets to search for a set of optimal parameters, thus enhancing the performance of the real–time leak detection program. The optimized parameters are the feature pool for training the k–NN classification model and the number k (the number of nearest neighbors used for the k–NN classifier). We perform the offline analysis process using a number of AE datasets, as shown in Table 4.2.

Table 4. 2. Number of datasets used for the offline analysis and evaluation.

	P$_0$		**P$_1$**		**P$_2$**	
	N_{FA}	**N_{FE}**	**N_{FA}**	**N_{FE}**	**N_{FA}**	**N_{FE}**
L$_0$	600	30,000	600	30,000	600	30,000
L$_1$	200	10,000	200	10,000	200	10,000
L$_2$	200	10,000	200	10,000	200	10,000
L$_3$	200	10,000	200	10,000	200	10,000

For feature selection, we should first normalize extracted features to place them on the same unit basis. The feature normalization is expressed by the following equation:

$$y_{\text{new}} = \frac{y_{\text{old}} - \mu_{yn}}{\sigma_{yn}} \tag{4.5}$$

where y_{old}, y_{new} are original and rescaled features, respectively, and μ_{yn}, σ_{yn} are successively mean and standard deviation of the feature estimated from samples belonging to the normal pipeline state.

Table 4.3 exhibits feature scores using the KL distance method. The most highly ranked features are STE, RMS, AVA, and STD, and these are returned in every pressure condition. Hence, we only consider these kinds of features to build the real–time gas leak detection program. Figure 4.5 illustrates the 3–D visualization of three features with the highest scores under diverse pressure conditions, in which the normal/leaking states are obviously separated for all the cases. Moreover, we know that a large k may improve performance; however, a too large k destroys the locality. Therefore, to choose k appropriately, we employ the available k–NN fitting function "*fitck–NN*" supported by Matlab 2019a to trial different values of k using the analysis datasets and we obtain $k = 25$.

Table 4. 3. Feature score based on KL distance.

	STE	**RMS**	**AVA**	**MEA**	**STD**	**ZCR**	**ETY**	**KUS**	**SKE**	**SPP**	**SPC**	**SPS**
P$_0$	57.7	36.5	37.3	−30.2	36.2	10.8	4.0	6.9	−18.0	1.8	7.0	8.9
P$_1$	71.9	44.2	44.4	−0.2	44.0	10.2	−1.7	3.3	−9.1	2.0	7.8	7.6
P$_2$	77.7	47.2	47.4	5.1	47.1	13.3	−5.6	1.1	−6.5	3.3	11.1	9.0

The datasets belong to a signal channel (R15i Ch1 or R15i Ch2), corresponding to three pressure conditions: 700 kPa (P_0), 1300 kPa (P_1), and 1800 kPa (P_2), and pipeline health states: normal (L_0), leaking (0.3 mm (L_1), 0.5 mm (L_2), and 1 mm (L_3)), which were recorded in Subsection 4.2; N_{FA} and N_{FE} are the numbers of frames for the offline analysis and experiment respectively, and a frame consists of 8192 samples stored in the hard drive.
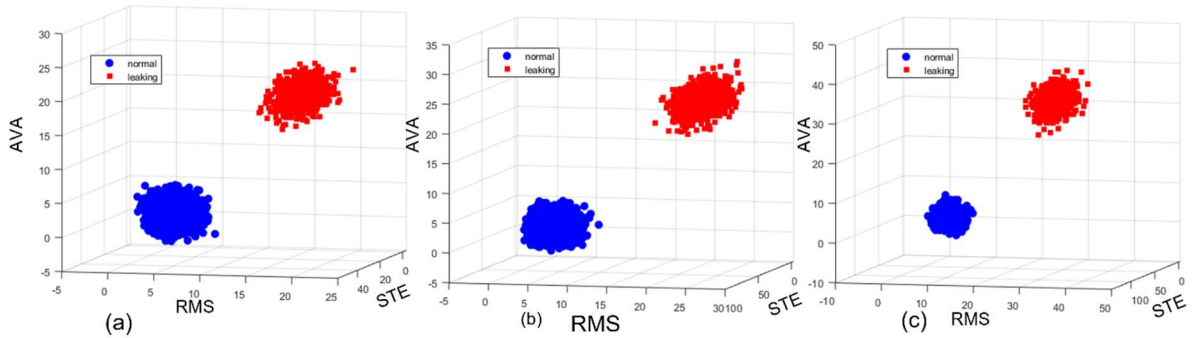
Figure 4. 5. Three-dimensional visualization of the three most highly ranked features under various pressure conditions: (a) P_0, (b) P_1, (c) P_2.

4.4.2. Gas Pipeline Leak Detection Implementation on an MCU-Based Hardware Architecture

(1) Overview of the Experimental Hardware Design with an MCU Used for Real–time Gas Pipeline Leak Detection
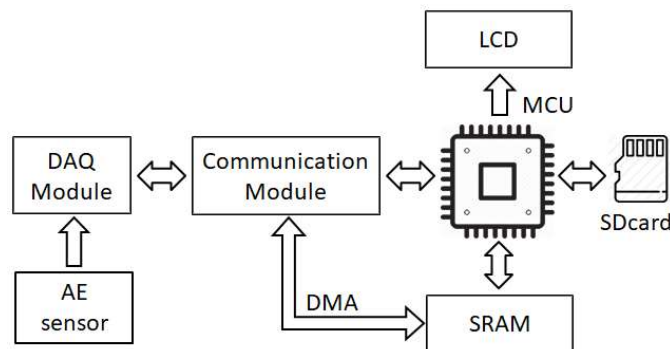
Figure 4. 6. Experimental MCU–based hardware architecture for the gas pipeline leak detection.
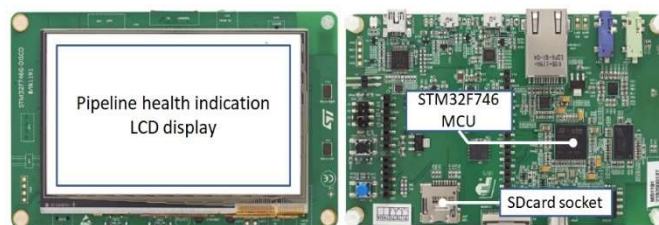
Figure 4. 7. 32F746G-DISCOVERY board (top view: left side—bottom view: right side).

Figure 4.6 illustrates an MCU–based hardware architecture to implement the proposed method for real–time gas pipeline leak detection. A sensor channel is connected to a DAQ module

which converts analog AE signals to digital AE signals and directly writes them to a SRAM through a communication module, along with a DMA channel available in the MCU; hence, the leak detection program can investigate AE signals in real–time. We also design a portable memory (SDcard) to store some pre–defined parameters of the leak detection program and its runtime log files used for later analyses. Hence, the program can be adjusted and updated quickly. Additionally, a LCD is installed to indicate the output of the diagnostic program. This entire design is embedded in the 32F746G–DISCOVERY board, as shown in Figure 4.7.

(2) Real–time Gas Leak Detection Implementation on the 32F746G–DISCOVERY Board

```
/* STE feature */              /* AVA feature */              /* Normalization */
STE = 0;                       AVA = 0;                       y =[STE, AVA, RMS, STD]; // a feature vector
for (i = 0; i < N; i++)        for (i = 0; i < N; i++)        y = floor (10*(y-μyn)/σyn);
    STE += (x[i] * x[i]);          AVA += abs(x[i]);          /* k-NN classification */
end                            end                            // Manhattan distance calculation
                               AVA = (AVA / N);               dist = dcal_func (v_train, M, y);
/* STD feature */                                             // Search for k minimum Manhattan distances
a = mean (x, N);               / * RMS feature */             NN_index = find_nn_func(dist, k);
STD = 0;                       RMS = 0;                       // Predict class
for (i = 0; i < N; i++)        for (i = 0; i < N; i++)        [n_NORM, n_LEAK] = count_func(label, NN_index);
    STD += (x[i] - a) * (x[i] - a);    RMS += (x[i] * x[i]);  if (n_NORM > n_LEAK) class = NORM;
end                            end                            else  class = LEAK; end
STD  = sqrt(STD / (N - 1));    RMS = sqrt(RMS / N);
```

/* x is an array of N raw AE signal samples; v_train is an array of M training feature vectors loaded from the sdcard; μyn and σyn are mean and standard deviation of normal state features, updated frequently in run-time */

Figure 4. 8. Primary program module of real–time gas pipeline leak detection embedded in the 32F746G–DISCOVERY board.

Due to the limit on memory and computing speed of an MCU, we use integer instead of floating–point format for the feature calculation and the k–NN classification, thus utilizing the memory economically and lightening the computation load. In other words, a real feature value is multiplied by 10 before rounding it, which sustains a one–decimal point precision for the vectors of rounded features, while avoiding reduction in the classification quality.

A trained classifier leans heavily on its training datasets, while AE signals acquired from a pipeline are prone to variation because the inner flow rate and pressure change constantly. The signals also fluctuate according to the sensor installation location and the operating moment. To reconcile these differing environments, we must adjust the trained leak detection model to its real and specific operational conditions. Therefore, the paper proposes updating the classifier by modifying the two parameters μ_{yn} and σ_{yn} related to the normal pipeline state in run–time, and which are employed in (4.5). Figure 4.8 shows the feature calculation and k–NN classification module of a real–time gas pipeline leak detection program implemented on the 32F746G–DISCOVERY board.

4.5. Experimental Results

To evaluate the gas pipeline leak detection system quickly, we emulate a real DAQ using a computer program which dispatches recorded AE signal datasets, whose description is shown in Table 4.2, through an available communication channel to the 32F746G–DISCOVERY board. This does not affect the objective assessment because the datasets have been acquired from a practical pipeline testbed under various operating conditions. We here figure out three aspects: detection accuracy, real–time characteristic, and detection robustness, because those are key factors to apply a leak detection system for the real environment.

4.5.1. Detection Accuracy and Real–time Characteristic
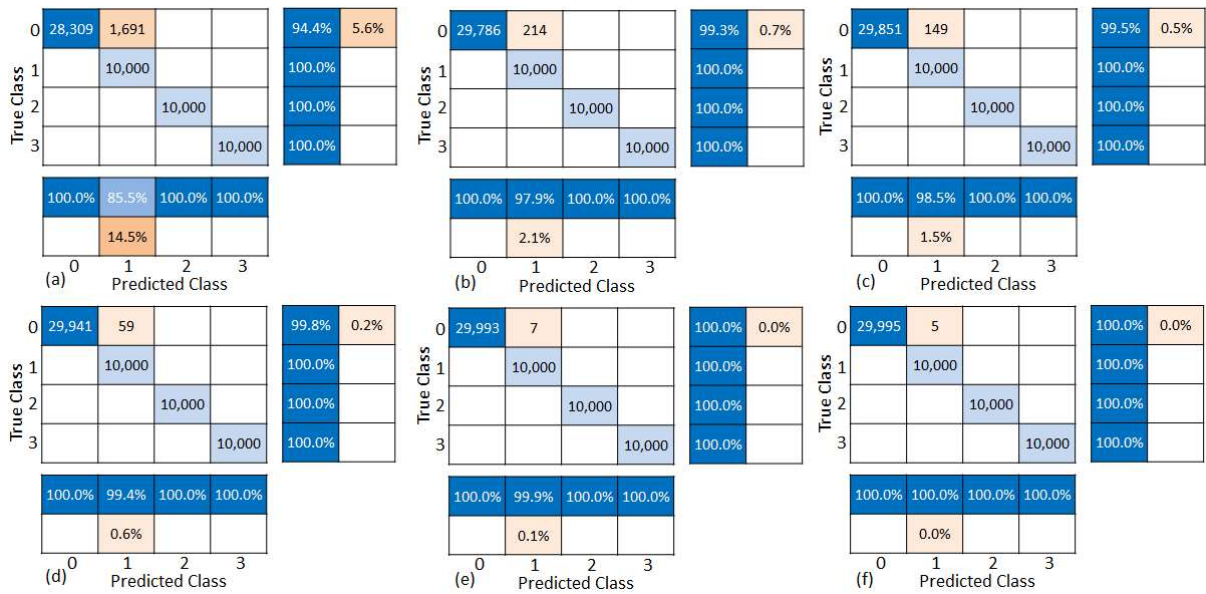
Figure 4. 9. Confusion matrices resulting from experimental scenarios: R15i Ch1 (a) P_0, (b) P_1, (c) P_2; R15i Ch2 (d) P_0, (e) P_1, (f) P_2 (classes 0, 1, 2, 3, and 4 are L_0, L_1, L_2, and L_3, respectively).
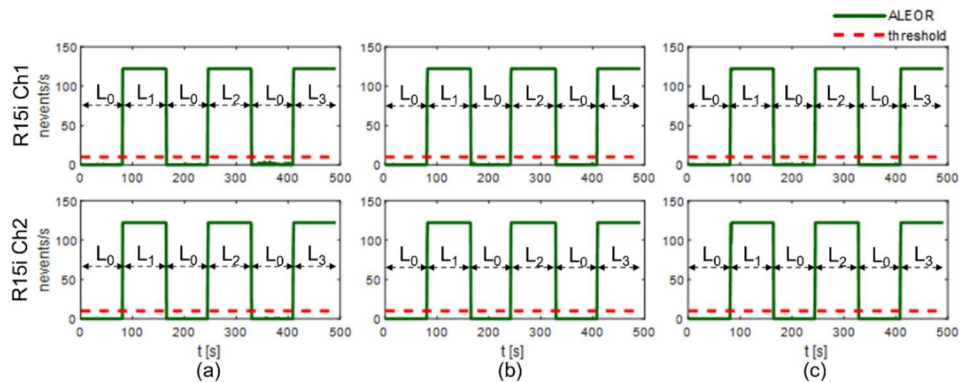
Figure 4. 10. ALEOR under different pressure conditions: (a) P_0, (b) P_1, (c) P_2.

Figure 4.9 shows confusion matrices of experimental results returned by the leak detection program running on the 32F746G–DISCOVERY board, and Table 4.4 illustrates classification accuracy and execution time for evaluation scenarios. The accuracy, as averaged over the two

sensor channels (R15i Ch1 and Ch2), and that of various pipeline states (L_0, L_1, L_2, and L_3), is relatively high at better than 98% for every pressure condition (P_0, P_1, and P_2). Besides, the mean execution time (t_E = 109 s) is less than the total experimental dataset duration (t_D = 123 s). This demonstrates the real–time characteristic of the implemented system that does not miss any data and returns a timely result during the analysis operation. Furthermore, the ALEOR is monitored while examining dataset pairs (L_0, L_1), (L_0, L_2), and (L_0, L_3) subsequently (see Figure 4.10). This plot reveals the correct identification of pipeline states: normal (L_0), leaking (L_1, L_2, and L_3), exploiting a threshold γ = 10 (see red dash line in Figure 4.10). The leaking state is decided only if ALEOR exceeds the threshold, despite the fluctuation below it. Therefore, no false alarm is reported in the experiment and the leaking state is also indicated punctually.

Table 4. 4. Classification accuracy and execution time.

		P_0			P_1			P_2		
		A	t_D	t_E	A	t_D	t_E	A	t_D	t_E
	L_0	97.2	246	214	99.7	246	214	99.8	246	214
R15i	L_1	92.8	82	74	99.0	82	74	99.3	82	74
Ch1	L_2	100	82	74	100	82	74	100	82	74
	L_3	100	82	74	100	82	74	100	82	74
	L_0	99.9	246	214	100	246	214	100	246	214
R15i	L_1	99.7	82	74	100	82	74	100	82	74
Ch2	L_2	100	82	74	100	82	74	100	82	74
	L_3	100	82	74	100	82	74	100	82	74
Average		98.7	123	109	99.8	123	109	99.9	123	109

where t_D and t_E are the total time of datasets and execution time, respectively, measured in seconds. A is classification accuracy given by: A = 100xN_C/N_{FE} [%], N_C is the number of correctly classified frames.

4.5.2. Detection Robustness

The result as exhibited in Table 4.4 and Figures 4.9 and 4.10 is obtained by using the test datasets under the same recording condition as the training datasets. As a result, the effectiveness of the proposed leak detection system may not be adequately demonstrated, because in a real gas pipeline network, there are always irregular disturbances leading to AE signal modifications, such as operating mode variation (inner pressure or flow rate), noise interference, etc. Measurement of an AE sensor can be modelled as follows:

$$z = x + \eta \tag{4. 6}$$

where z and x are measured and original signals, respectively, and η represents any signal modification including ambient noise and discrete events. We assume the normal distribution function for both x and η. According to the probability rule specified by [64], z distributes

normally also, and its mean and standard deviation are sequentially:

$$\mu_z = \mu_x + \mu_\eta; \quad \sigma_z = \sqrt{\sigma_x^2 + \sigma_\eta^2} \tag{4.7}$$

where μ_z, σ_z, μ_x, σ_x, μ_η, σ_η are means and standard deviations of z, x, and η, respectively. Equation (4.7) shows that the abnormal disturbance distorts original signals, thus deteriorating the signal-based leak detection model.
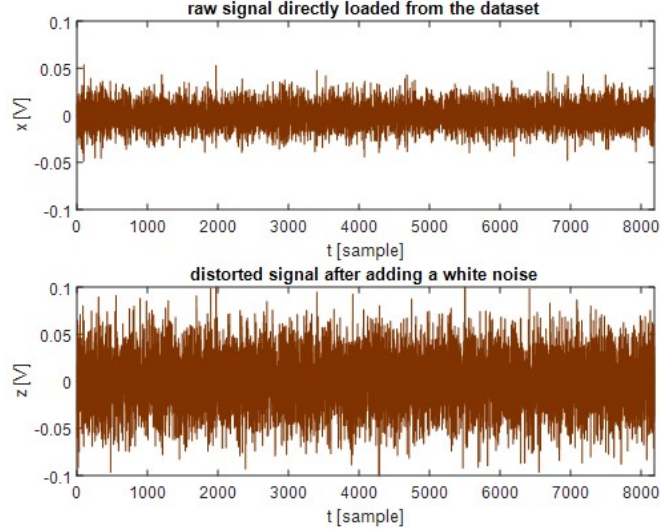
Figure 4. 11. A signal after adding a white noise with $\rho = 2$.

To verify the robustness of the proposed leak detection method, we add white noise to the experimental datasets prior to conducting the real–time leak detection on the 32F746G–DISCOVERY board. This noise is referred to as the signal disturbance η, simulated by an available function in the Matlab software with a rule below:

$$\mu_\eta = 0; \quad \sigma_\eta = \rho \times \sigma_{xn} \tag{4.8}$$

where σ_{xn} is the standard deviation of normal state signal (acquired when the pipeline is healthy), and ρ is a proportion ratio. We set $\mu_\eta = 0$ in (4.8) because the mean parameter of a signal is mainly related to low frequency components of that signal, while the operating frequency range of R15i sensors is from 50 kHz to 400 kHz. The low frequency band (below 50 kHz) is not examined and the influence of μ_η is therefore relatively minor or $\mu_\eta \approx 0$. Figure 4.11 illustrates the signal distortion if adding a white noise η according to (4.6) and (4.8) where $\rho = 2$. We can easily realize that the distorted signal energy is greater than the original because of the added noise in Figure 4.11.

We alter ρ and observe the performance deterioration of the trained classifier. Figure 4.12 shows the dependence of ROC and ACA on ρ. The computation is calculated on all the datasets of the

two sensor channels in two cases: with updating μ_{yn} and σ_{yn} (see Subsection 4.4.2 (2)) and without updating. The classification performance substantially declines at slight values of ρ if we do not adapt the model to the increasing added white noise (see Figure 4.12 (a) and the blue dash dot line in Figure 4.12 (c)). In contrast, the classifier can still work acceptably until $\rho = 70$ if we adjust μ_{yn} and σ_{yn} (see Figure 4.12 (b) and the red solid line in Figure 4.12 (c)). With $\rho = 10$, the resulting classification accuracy is above 90% (see Figure 4.12 (c)) and the pipeline state can be exactly identified by the ALEOR with a threshold $\gamma = 10$, as shown in Figure 4.13 for every experimental condition. In short, the proposed methodology can ensure the robustness of the leak detection system.
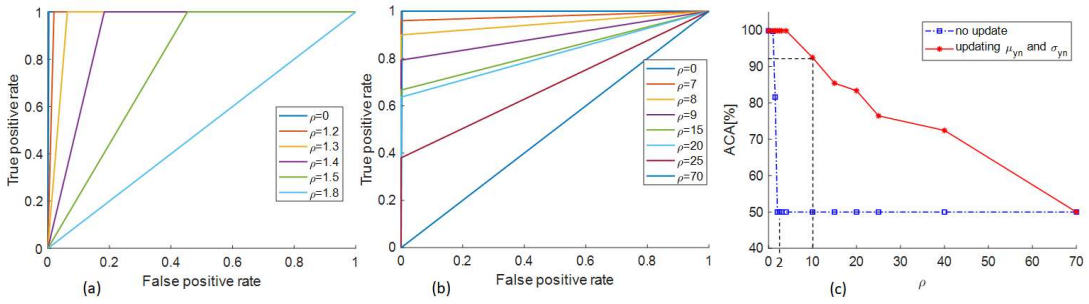
Figure 4. 12. ROC and ACA according to ρ: (a) ROC without updating μ_{yn} and σ_{yn}, (b) ROC with updating μ_{yn} and σ_{yn}, (c) ACA reduction.
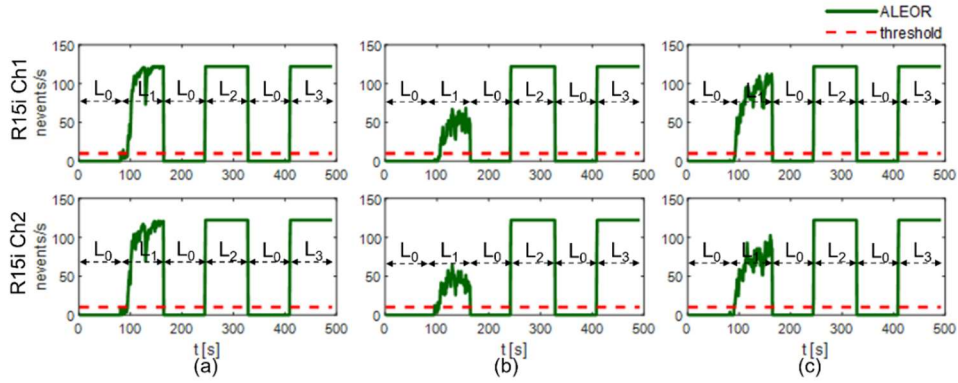
Figure 4. 13. ALEOR under different pressure conditions: (a) P_0, (b) P_1, (c) P_2 (After adding a white noise with $\rho = 10$).

Although the proposed method can sustain a high classification performance with small values of ρ, the classification performance still deteriorates gradually according to the increase in ρ and the classifier cannot precisely operate with $\rho > 70$ which causes severe distortion of the acquired signals. Therefore, we should configure the testbed to resemble an applied real pipeline before gathering datasets for training the classifier, thus obtaining an adequate leakage detector. The greater the similarity between the testbed and the real pipeline, the more accurate the detection is.

4.6. Conclusions

A complete system is offered for real–time gas pipeline leak detection in the paper. First, the system offline analyzed recorded AE signals sampled at 1 MHz. The process configured a hybrid feature pool and normalized its elements using the mean and standard deviation of the set of feature observations related to normal pipeline health. Then, the pool was purified using the three-sigma rule and the Kullback–Leibler distance to obtain the most discriminative signatures. Next, the system identified the pipeline health states (normal/leaking) with an input vector of features, by exploiting a k-nearest neighbor classifier that seeks the purified feature pool for the signatures closest to the input vector, based on the Manhattan distance. To avoid issuing a false alarm, the system decided a pipeline state via monitoring the accumulative leaking event occurrence rate and a predefined threshold. Finally, the total proposed leak detection method was embedded in a compact MCU–based hardware platform for real–time leak detection. The detection accuracy, the real–time characteristic, and the robustness of the introduced gas pipeline leak detection system have been evaluated. The experimental results showed that the system indicated pipeline health states robustly in a quick enough timeframe for real–time application. Thus, this system can be applied for inspecting pipeline health in a real gas pipeline network.

The testbed used in this paper for collecting AE signals is a part of a real gas pipeline network. Hence, the resulting AE signals are not simple signals generated by the pipeline leakage simulation in a laboratory. They do not only contain information about pipeline states (normal or leaking), but also depend on practical gas transportation and systematic behavior. Additionally, a noisy measurement location and wave attenuation could conceal symptoms of leakage in recorded signals. This challenges the signal investigation because the relation between the leakage phenomenon and AE signals is unclear in the initial analysis stages. Therefore, a short pipeline was chosen in the paper to easily separate signal classes related to pipeline states corresponding to different experimental scenarios, hence conveniently proposing a leak detection method as well as evaluating experimental results. However, it is believed that the proposed technique can effectively monitor a long pipeline in a real application. The pipeline length depends on the signal detection ability of the AE sensor—their sensitivity and a specific working environment. These parameters can be estimated by using pencil lead breaking tests.

Chapter 5

Crack Detection and Localization in a Fluid Pipeline using Acoustic Emission Signals

5.1. Introduction

Aside from assembly defects, leakage in a pipeline network generally originates from ruptures resulting from material aging and from the impact of inner pressure. This process usually happens in a period referred to as crack growth. Initially, a crack may be a small structural deformation when a pipeline is under pressure slightly higher than its current endurance. The flaw may gradually become a fracture if the pipeline continues to operate. Therefore, a pipeline health inspection program is essential in providing the ability to alert operators of any irregularities in pipeline health early and thus avoid leakage. During the crack growth period, AE signals are generated by the stress and strain at the deformation position [6, 11, 65, 66]. Many studies offer algorithms for analyzing such signals to detect and localize cracks in material structures [33, 67, 68]. Those install AE sensors on the surface of the experimental object to collect AE signals (known as AE waves) that are propagating throughout the body and surface of the object from deformed points. The AE signals contain relevant information about a probable crack, even at its inception; as a result, the AE based approach can immediately detect any small variation in the material structure.

Although AE signals directly respond to an imminent fracture, clearly identifying unusual syndromes in these signals is challenging, due to noise presence. This is definitely the case for fluid pipelines because the inner medium flow makes a nonstationary noise. Hence, an adequate algorithm should be developed to extract fault symptoms from AE signals. Currently, there is a study using AE parameters such as rise time, duration, amplitude, ring down count, and energy in the time domain. This study constructs the event probability by relying on parameter variations versus loading cycles [33]. In this method, a cumulative AE event count and event rate versus time may indicate a fatigue crack. However, this method does not exclude noise events from the computation, and noise is abundant in such cases. Therefore, this method might be inapplicable for a real pipeline network (for example, AE signals which are measured at a pipeline location in a factory could be inevitably influenced by nearby elements and fluid flow fluctuation due to complexity of pipeline topology). There are several approaches that detect cracks in concrete structures through a trained model [18, 69] with high accuracy. However, a key requirement of these approaches is that training data must be available. This requirement may not be easily satisfied with pipelines, in which real signals vary with operating conditions

such as temperature and inner pressure [11, 21, 42]. Moreover, acquired signals are generated by various sources in a pipeline system (for example, nearby machinery vibrations, and inner flow turbulences). As a result, dataset labeling that categorizes data into classes for learning is problematic. Another AE–based technique (used for crack detection in pressure vessels and tanks) determines hits in the time domain by means of a threshold, and clusters AE sources according to their coordinate and emission time [70]. The paper states that a region with a high density of hits could be a sign of impending fracture. This statement is reasonable; however, the location accuracy significantly affects hit clustering, while hit localization still results in significant errors, as specified in Ref. [70]. There are two main reasons for such location errors. First, diverse wave modes with different velocities could be stimulated, and they could convert into or interfere with each other [11]. Second, AE waves attenuate along their propagation path, depending on not only distance, but also the frequency [37, 39, 71]. Some works propose solutions for enhancing localization accuracy via separating wave modes and improving onset time determination [72-81] in time domain signals. However, the nonstationary noise in operating pipelines would substantially influence those procedures, and the location accuracy could still be inadequate for source clustering based on their coordinates.

As analyzed above, for crack detection and localization, current AE based methods would be ineffective if used for fluid pipelines. Hence, this paper proposes a novel approach, incorporating existing ideas along with improvements for early detection and localization of cracks in a fluid pipeline. Our approach detects AE events in time-frequency domain signals instead of time domain signals. This provides a frequency feature to eliminate unwanted emission sources while localizing AE sources, thus increasing the location accuracy. To reduce the coordinate errors further, the algorithm separates wave modes to identify Rayleigh waves for matching wave velocity with flight time. Hence, the idea of clustering sources according to their coordinates can be effectively exploited because the locations have become accurate. Furthermore, this paper constructs a line of AE activity against applied load using filtered AE events. The variation in AE activity can reveal irregular states in the pipeline structure, which is claimed by the theorem of Kaiser and Felicity effects [11, 82]. In addition, an AE source distribution density in terms of the source location coordinates along the pipeline (AE source histogram) can specify an abnormal region where there is a high source concentration. The experimental results show that the proposed method obtains high accuracy in localizing AE sources, thus facilitating early crack detection.

5.2. Data Acquisition
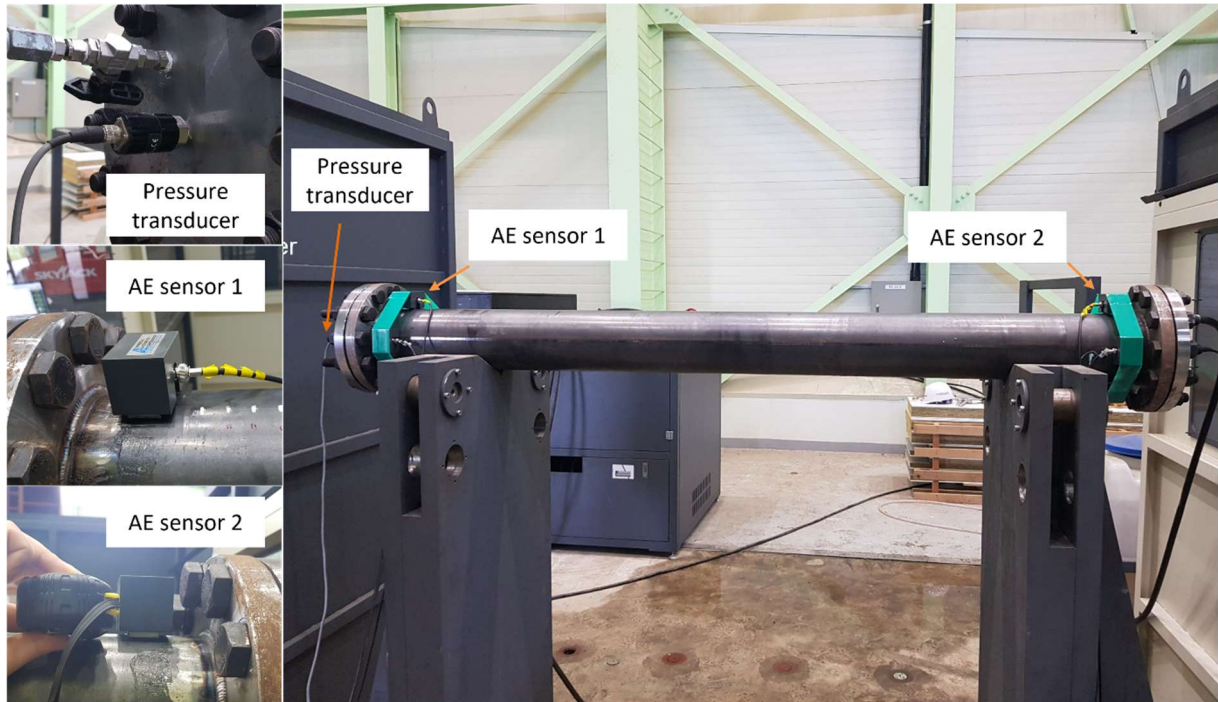
Figure 5. 1. Pipeline testbed for crack detection

Figure 5.1 illustrates a pipeline testbed that was constructed to record AE signals for crack detection, and Figure 5.2 depicts its configuration. We selected a pipeline specimen made from carbon steel, and water as the transported medium in our experiment. The pipeline's geometrical parameters are as follows: 2 m length, 4.85 mm thickness, and 165.2 mm outer diameter. Water flow was circulated from a water tank with a pump and external specialized pipelines, and the inner pressure was controlled by two valves, as shown in Figure 5.2 (a). A pressure transducer was installed in the system to monitor any pressure change.

In accordance with the BSI standard BS EN 15856 [42], the relevant frequencies used for analyzing AE wave propagation in metal range from 100 to 300 kHz. Therefore, we set up two R15i–AST sensors at the two ends of the pipe to acquire AE signals while pressurizing the pipeline. This type of sensor has a resonant frequency of 150 kHz, and a range of operating frequencies from 50 to 400 kHz [41], which covers the above frequency range. To store signals in a hard drive, we utilized an integrated DAQ system based on an EXPRESS–8 device designed by MISTRAS Group [83] (see Figure 5.2). We digitalized AE signals at a sampling rate of 1 MHz, which was chosen according to the Nyquist–Shannon sampling theorem, in which the sampling frequency must be at least double the maximum frequency of the signal, as stated in [63]. After configuring the hardware and software correctly, we exploited the pencil lead breaking technique [58] to check the sensitivity of the mounted sensors and the entire operation of the experimental system. Hence, the recorded dataset is reliable for later analyses.
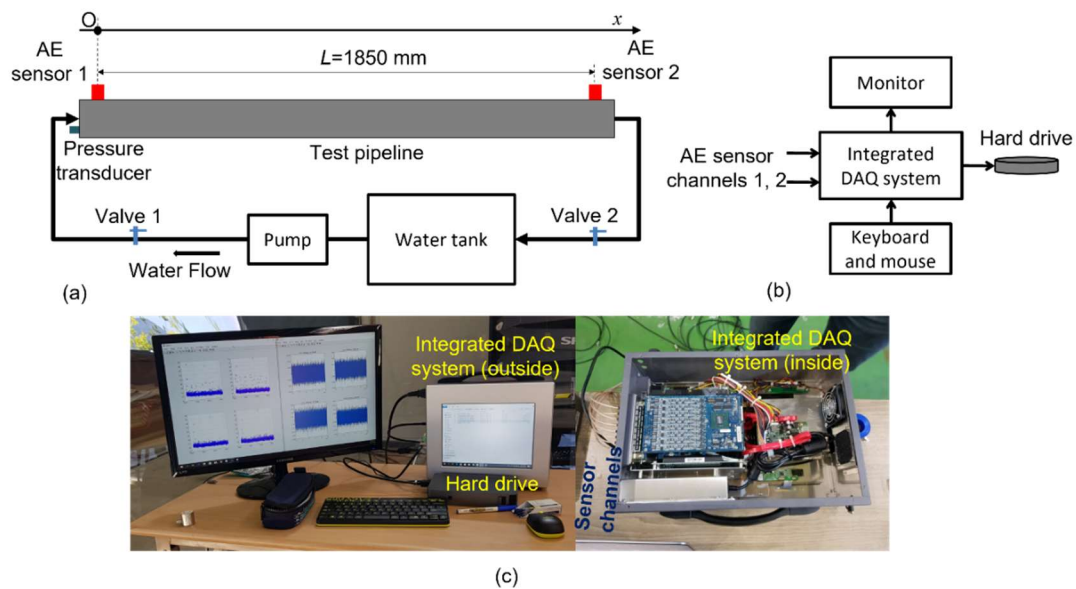
Figure 5. 2. Experimental configuration and data acquisition system: (a) testbed, (b and c) integrated DAQ
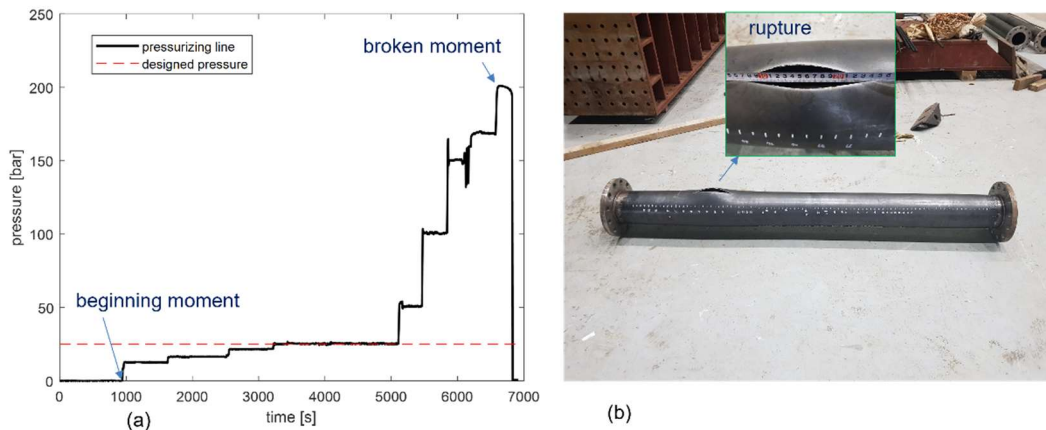
Figure 5. 3. (a) Pressurizing process, (b) the broken pipeline

To observe the variation in AE signals resulting from a crack, we pressurized the pipeline sample since it was intact when the inner pressure was 0 bar, until it ruptured. Figure 5.3 (a) shows the increasing pressure line and Figure 5.3 (b) shows the fractured pipe detached from the testbed. In a real situation, we would never increase the pressure beyond the design pressure of the pipeline (indicated by the red dash line in Figure 5.3 (a)), because of safety concerns. However, we conducted the destructive experiment to correlate the crack phenomenon happening in a water pipeline under high pressure with AE activity. Moreover, operating pipelines gradually degrade over time, which results in a lower critical pressure than their nominal design pressure. Thus, depending on the level of pipeline degradation, over time, a serious fault may occur, even if the operating pressure is less than the design pressure. Hence, the experiment emulated what might happen in real applications except for using high pressure to damage the pipeline.

5.3. Methodology

The crack detection and localization algorithm for a fluid pipeline is illustrated in Figure 5.4. The methodology consists of three procedures: event detection, source localization, and source monitoring. AE signals are acquired by the corresponding AE sensors and saved in the hard drive, as shown in the data acquisition subsection, and then are loaded as inputs of the diagram. Figure 5.4 shows an offline analysis. However, we can apply this algorithm online for real-time situations in which the data acquisition and the pipeline health inspection are simultaneous.
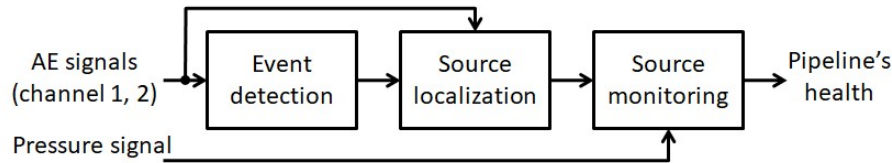
Figure 5. 4. The overall diagram of crack detection and localization for a fluid pipeline

5.3.1. AE event detection

The method using AE hits to identify suspicious faulty regions detects events directly through bursts of AE signals in the time domain [34]. As stated in the introduction, this could be productive if noise is stationary while background noise of AE signals measured from an operating pipeline is non–stationary, because it is extremely influenced by dynamics of its inner water flow. Furthermore, the wave attenuation weakens bursts in signals acquired by sensors when they propagate from a distant source to measurement places, thus challenging the event detection in the time domain. In order to address this problem, we investigated events in the time–frequency domain, as illustrated in Figure 5.5. Despite the amplitude attenuation in the time domain, an AE event can still emerge from the noise background, due to its distinct frequency content. Additionally, a detected event in the time–frequency domain contains not only its time of arrival and amplitude (as in the time domain) but also its frequency. This information is helpful in pairing events correctly during the next stage, in which we localize their source (see Subsection 5.3.2).
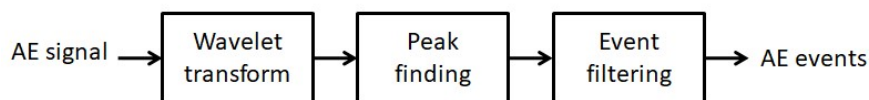
Figure 5. 5. The event detection diagram

The Wavelet transform can perform a detailed time–frequency analysis of a non–stationary signal better than the Fourier transform or the short time Fourier transform [84]. Thus, we construct the Wavelet transform in the first block of the event detection diagram to decompose a one–dimensional time domain signal into a two–dimensional time–frequency domain

Wavelet coefficient matrix where each row corresponds to one frequency band and the column size is equal to the length of signal in the time domain. The number of input samples and frequency bands must not too large to ensure the fast transform execution and not too small to obtain an adequate time–frequency content.

The Wavelet coefficient matrix is directly led to the peak–finding block. This is a complex matrix containing both the phase and amplitude; however, we only analyze its amplitude components to seek local maxima, which are candidates for events in the AE signal. Since the two–dimensional amplitude matrix is similar to a gray image (a pixel is an element with coordinate of frequency band and sample index), the technique seeking local maxima in a gray image [85] can address our peak–finding problem. The overall peak finding algorithm is shown in Figure 5.6.
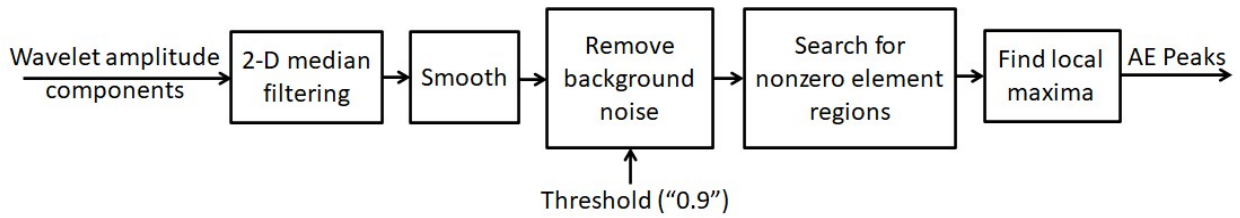
Figure 5. 6. Peak finding algorithm

First, a 2–D median filter is utilized to get rid of single pixel noise [86]. Second, we smooth the filtered matrix so that it exhibits a high probability for only one pixel in each peak corresponding to a local maximum. We continue removing the noise background of the pixel matrix using a threshold as follows:

$$a_{ij} = \begin{cases} 0 & if \quad \dfrac{a_{ij}}{a_M} < \alpha \\ a_{ij} & if \quad \dfrac{a_{ij}}{a_M} \geq \alpha \end{cases} \tag{5. 1}$$

where a_{ij} is the (i^{th}, j^{th}) element of the pixel matrix, a_M is the maximum pixel value of the pixel matrix, and α is the threshold. In fact, the threshold can be adjusted, however, we set a default value $\alpha = 0.9$ for the quick peak finding. Finally, we scan pixel regions containing nonzero elements in the processed pixel matrix and search for local maxima from those regions to obtain AE peaks (including their frequency band, sample index, and amplitude). The result of peak finding is shown in Figure 5.7, in which local maxima are marked by red circles.

The detected peaks could come from noise not due to AE events because AE activity related to material structure is present only if the load levels exerted on the material exceed the previous

stress (Kaiser effect) according to Ref. [11]. Therefore, a filtering step is necessary to obtain true events. We establish the filter based on a Neyman–Pearson theorem of signal detection probability, with an assumed normal distribution for the noise [57], which is illustrated by Eq. (5. 2).

$$L(z) = \frac{p(z \mid H_1)}{p(z \mid H_0)} > \gamma, \quad P_{FA} = \int_{\{z:L(z)>\gamma\}} p(z \mid H_0) dz \tag{5. 2}$$

where $L(z)$ is the likelihood ratio, H_0 is the signal absence hypothesis, H_1 is the signal presence hypothesis, z is an observed set, $p(z)$ is the probability density function, P_{FA} is the false alarm probability, and γ is a threshold. In our study, H_0 and H_1 are hypotheses of the peaks coming from noise and events, respectively, and the likelihood ratio $L(z)$ is calculated according to the peak amplitudes.
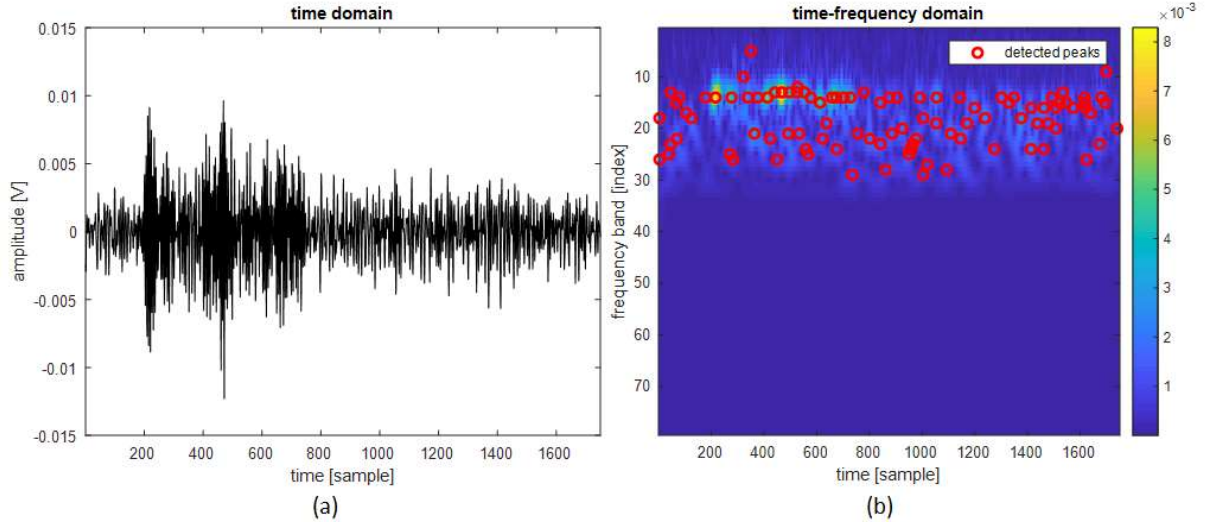
Figure 5. 7. (a) AE signal in time domain, (b) detected peaks in time-frequency domain (red circles)

Eq. (5. 2) is a general equation to calculate a threshold γ for classifying two hypotheses H_0 and H_1 (true events) with a false alarm probability P_{FA}. Since a normal distribution is assumed for the noise, the probability density functions $p(z|H_0)$ and $p(z|H_1)$ are as follows:

$$p(z|H_0) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}; \quad p(z|H_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu-x_0}{\sigma}\right)^2} \tag{5. 3}$$

where μ and σ are the mean and deviation of noise, respectively, and x_0 is a shifting amount of probability density function due to the true event presence ($x_0 > 0$). Hence, the likelihood ratio is given by (5.4).

$$L(z) = \frac{p(z|H_1)}{p(z|H_0)} = e^{-\frac{x_0^2}{2\sigma^2}} e^{\frac{(z-\mu)x_0}{\sigma^2}} \tag{5.4}$$

Eq. (5. 4) shows a monotonically increasing exponential function $L(z)$ according to z because $x_0 > 0$, thus the inequation $L(z) > \gamma$ belonging to Eq. (5. 2) is equivalent to the inequation $z > z_0$ where $L(z_0) = \gamma$. Accordingly, we directly use z_0 as a detection threshold in the observation zone (z) instead of using the threshold against the likelihood function $L(z)$ to simplify the computation. In other words, we set $u = (z-\mu)/\sigma$, $u_0=(z_0-\mu)/\sigma$, and the false alarm probability becomes:

$$P_{FA} = \int_{\{z:L(z)>\gamma\}} p(z|H_0)dz = \int_{u_0}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = \Phi(u_0) \tag{5.5}$$

Therefore, if a false alarm probability P_{FA} is given, we can determine u0 through Eq. (5. 5), and employ an equivalent detection criterion for identifying true events from AE peaks:

$$\frac{z-\mu}{\sigma} > u_0 \tag{5.6}$$

where if a peak's amplitude satisfies Eq. (5. 6), the peak is assigned to be a true event. In addition, we can nearly estimate the parameters μ and σ from a set $z = z_0, z_1, \ldots, z_{N-1}$ by exploiting following expressions:

$$\mu = \frac{1}{N}\sum_{n=0}^{N-1} z_n, \quad \sigma = \sqrt{\frac{1}{N-1}\sum_{n=0}^{N-1}(z_n-\mu)^2} \tag{5.7}$$

Eq. (5. 5), Eq. (5. 6), and Eq. (5. 7) are applied for the event filtering in this paper, in which a false alarm probability is given, and the other parameters are calculated according to it. The false alarm probability can be adjusted to the number of returned events in a real application. A huge number of events would challenge the computation and visualization, we should therefore decrease the false alarm probability to eliminate more noise peaks whereas we can increase the false alarm probability to obtain more events.

Figure 5.8 illustrates the filtering out of noise peaks. The black dashed line in Figure 5.8 (a) is the detection threshold. Peaks whose amplitudes are below the threshold line are noise and the remaining peaks are true AE events. Figure 5.8 (b) designates events and noise peaks with red triangles and green circles, respectively, and Figure 5.8 (c) shows events and noise peaks in the time–frequency domain.
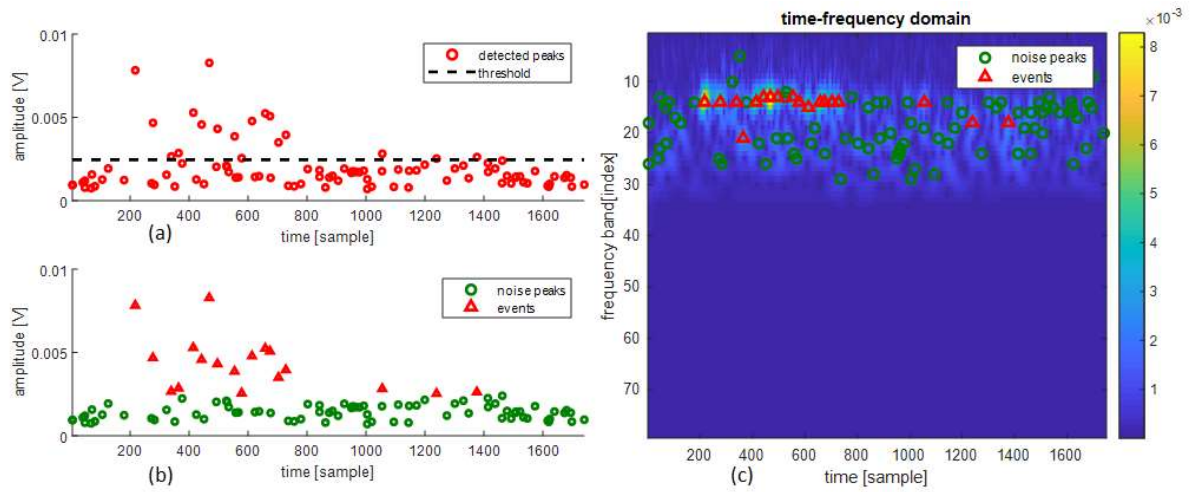
Figure 5. 8. AE event filtering: (a) detected peaks, (b) filtered events, (c) filtered events in time-frequency domain

In this manner, we can detect events in an AE signal in the time–frequency domain. However, one notable point is that removed peaks (which are considered as noise) could possibly come from faraway event sources whose signals have been seriously distorted. Therefore, we should choose an appropriate false alarm probability in order to not miss true events while eliminating unnecessary peaks. Inversely, events or peaks that pass the filter could possibly be false events. Their amplitudes could be great because their emission sources are intense and near sensors; however, they are out of the testing pipeline. The following stages will further eradicate false events, based on information about source location.

5.3.2. AE source localization

We establish a flow diagram for locating an AE source, as shown in Figure 5.9. Here, the core of localization exploits the TDOA between two signals, which has been thoroughly written about in Ref. [11]. The following mathematical equation demonstrates the technique:

$$x = \frac{L - C\Delta t}{2} \tag{5. 8}$$

where x is the AE source location, L is the distance between two sensors, C is the wave velocity, and Δt is the TDOA.

Eq. (5. 8) shows that the location accuracy depends on not only Δt, but also C. A mechanical wave propagates throughout a material in various modes, and their velocities are distinctive. Therefore, if we do not ascertain the wave mode, the value of C will be inappropriate to use in computing the location using Eq. (5. 8). In addition, different waves themselves tend to separate from each other in the dispersion from their emission source to sensors due to their distinctive

flight velocity, thus complicating the onset time determination. As a result, we should select the most distinct mode for the emission source localization. The pencil lead breaking investigation can expose proper modes in our specific pipeline. Figure 5.10 demonstrates a signal analysis when conducting a pencil lead breaking action at a coordinate of 0.6 m along the pipeline surface, according to the Ox axis shown in Figure 5.2 (a). In the experiment, we fully filled the pipeline with water at a pressure of 0 bar to take the pipeline and fluid interaction into consideration. As shown in Figure 5.10, three wave modes S_0, A_0, and Rayleigh can be identified while others are overshadowed. Consequently, detected AE events can fall in one of three modes: S_0, A_0, and Rayleigh. It can be seen that Rayleigh waves comprise most of the total energy. This result has been noted in other studies as well [20, 87]. Moreover, the noise background fluctuates when pressurizing pipelines; thus, the Rayleigh mode would be the most likely to indicate an event source localization because low-energy wave modes could be veiled by noise. A signal analysis performed when pressurizing the pipeline (see Figure 5.11) demonstrated this to be the case, in which the Rayleigh mode is the most dominant in both sensor channels.
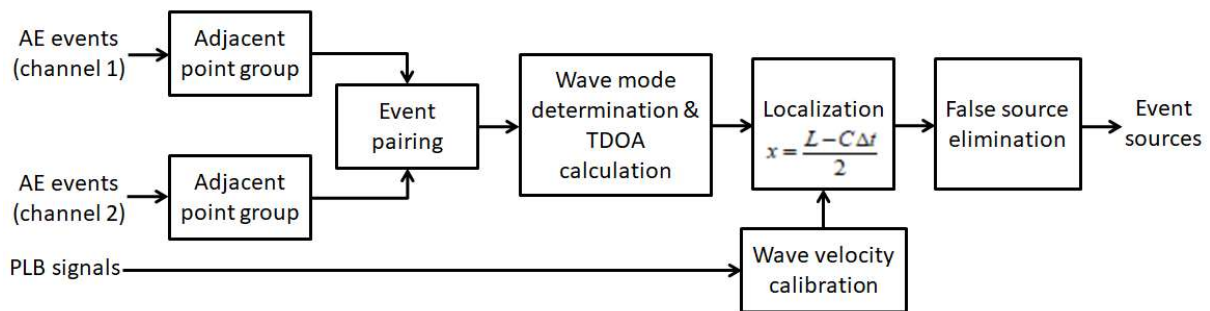
Figure 5. 9. AE source localization

The AE source localization algorithm shown in Figure 5.9 is carried out by using the Rayleigh wave mode. In the diagram, we first group neighboring events, before transferring them to the event pairing block, because an AE source results in ample adjacent events (see the filled red triangles in Figure 5.11). We call event collections "grouped events." They are marked with green–dashed ellipses in Figure 5.11. To group events, we scan all the detected individual events and examine their relation of sample index and frequency band. Two parameters ΔT and ΔF successively defined as sample index and frequency band distances are exploited to claim the neighboring relationship of two adjacent events. If $\Delta T \leq \Delta T_{\min}$ and $\Delta F \leq \Delta F_{\min}$, the two events are gathered in a grouped event, where $\Delta T_{\min}$ and $\Delta F_{\min}$ are minimum sample index and frequency band distances between the two individual events which are not their own neighbor. The values $\Delta T_{\min}$ and $\Delta F_{\min}$ can be flexibly modified in a real application. Additionally, we can further remove noise events by considering how many child events a grouped event contains.

This is illustrated by the following expression:

$$N_g \geq N_m \tag{5.9}$$

where N_g is a count of child events of a grouped event, and N_m is the minimum number.
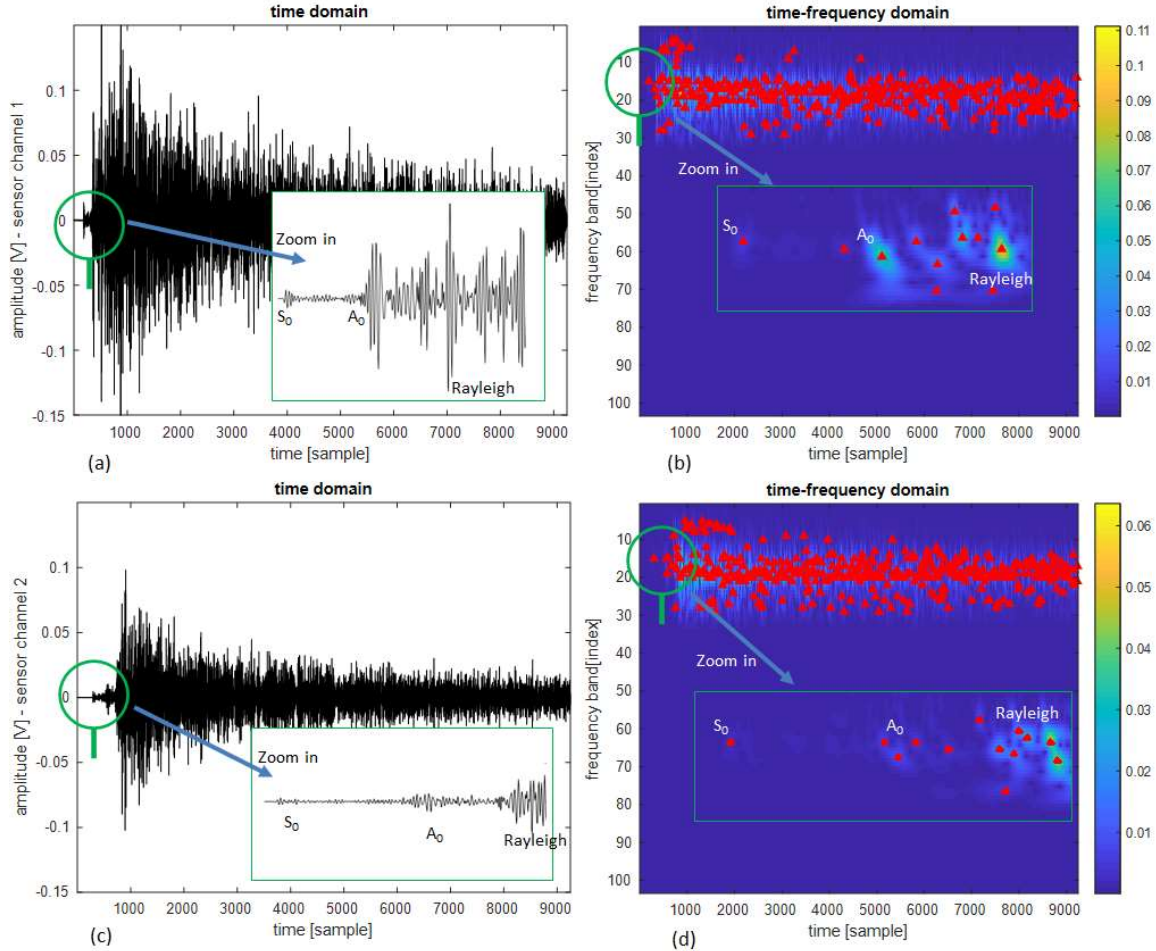
Figure 5. 10. Analyzing pencil lead breaking signals stimulated at the location of x = 0.6 m acquired by two sensors: (a, c) time domain signals, (b, d) AE events in time-frequency domain

Next, two grouped events are paired from two sensor channels for source localization, based on the time difference of arrival technique. With values $t_{1,2}$ (which are the onset time of grouped events of signal channels 1 and 2, respectively), we can simply compute source location via Eq. (8) and $\Delta t = t_2 - t_1$. However, to obtain an accurate location, we must precisely determine the Rayleigh mode to select $t_{1,2}$, and they should correspond to events with the same frequency band regarding both signal channels. Since the Rayleigh mode is the most dominant, it can be identified correctly according to the energy distribution. The onset of Rayleigh wave is located at the frontal first position where its energy is equal to 30% of the maximum energy region of a grouped event regardless of the other wave presence.
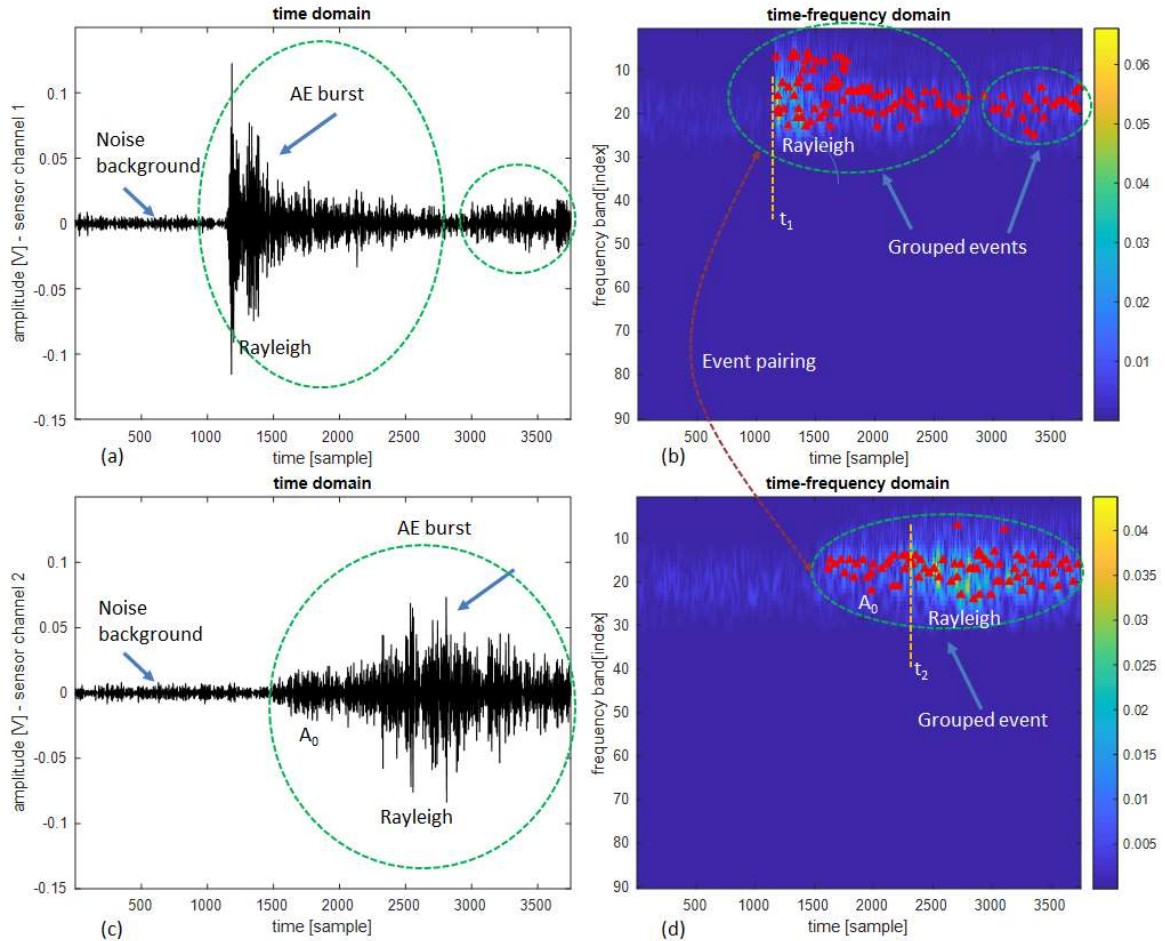
Figure 5. 11. Analyzing real signals acquired by two sensors when pressurizing a pipeline: (a, c) time domain signals, (b, d) AE events in time–frequency domain

Apart from errors returned by the wave mode determination and TDOA calculation, wrong event pairs also result in false alarms. For example, Figure 5.11 (b) comprises two grouped events of signal channel 1. However, the earlier one is actually matched with the grouped event of signal channel 2, as shown in Figure 5.11 (d). Thus, a step of false source elimination should be added. This is carried out by the following condition:

$$(A_1 - A_2)(t_1 - t_2) \le 0 \tag{5.10}$$

where $A_{1,2}$ are amplitudes of grouped events of signal channels 1 and 2, respectively, which are averaged over child event amplitudes. Eq. (5. 10) is determined by attenuation characteristics when AE waves propagate faraway [37]. This means that a sensor farther away receives less wave energy than a sensor that is nearby.

5.3.3. AE source monitoring

We monitor the AE activity while pressurizing a pipeline (see Figure 5.12 (a)) via visualizing AE sources according to their location, onset time, and amplitude, as shown in Figure 5.12 (b).

The plot draws the coordinates of AE sources symbolized by filled, colored circles. The size and color of the filled circles are linked to emission intensity (magnitude and duration). The source magnitude and duration are related to the average amplitude, as well as the mean number of elements of the grouped events that we paired to localize a source as presented in subsection 5.3.2. In this picture, a bright red circle with a large diameter depicts a strong source, and dark circles with small diameters indicate weak ones. Hence, we can conveniently observe the occurrence of AE event sources on the testing pipeline while increasing its inner pressure.

Next, we establish an AE activity vs. applied pressure graph, as depicted in Figure 5.12 (c). The AE activity–pressure graph can indicate the structural integrity of the pipeline material. This procedure refers to both the Kaiser effect theorem, describing an AE pattern in the material that is subjected to mechanical stress, in which AE events occur until the load is exceeded, and the Felicity effect, in which emissions continue to occur while the load is held [11]. Hence, if a load was kept still, yet the AE activity had not stopped or did not gradually vary due to noise impact, then there would be a major structural defect in the pipeline. For instance, the circled points in Figure 5.12 (c) are the "suspicious" signs, indicating risk of serious deformations in the pipeline structure because the AE activity sharply grew, even though we did not strengthen the pressure load.

The existence of energetic sources could indicate unsafe conditions in the structure of the pipeline. However, minor sources could be created by slight stresses and strains occurring in initial structural deformation stages, and these would also provide helpful information about possible crack locations. We therefore construct a histogram as illustrated in Figure 5.12 (d) by computing the distribution density of all the AE sources according to their coordinates along a pipeline. The histogram is established by equivalently dividing the pipeline into n small segments whose length is given by:

$$b = \frac{L}{n} \tag{5.11}$$

where b and L are the length of a segment and the whole pipeline, respectively. Besides, we usually predefine a relative location error er as follows:

$$e_r = 100 \times \frac{\Delta x}{L} \tag{5.12}$$

where Δx is the absolute location error.

To guarantee the coordinate discrimination for the histogram, we have (5. 13).

$$b \leq \frac{\Delta x}{2} \tag{5. 13}$$

From Eq. (5. 11), Eq. (5. 12) and Eq. (5. 13) we can reasonably choose the parameter *n* using the following constraint:

$$n \geq \frac{200}{e_r} \tag{5. 14}$$
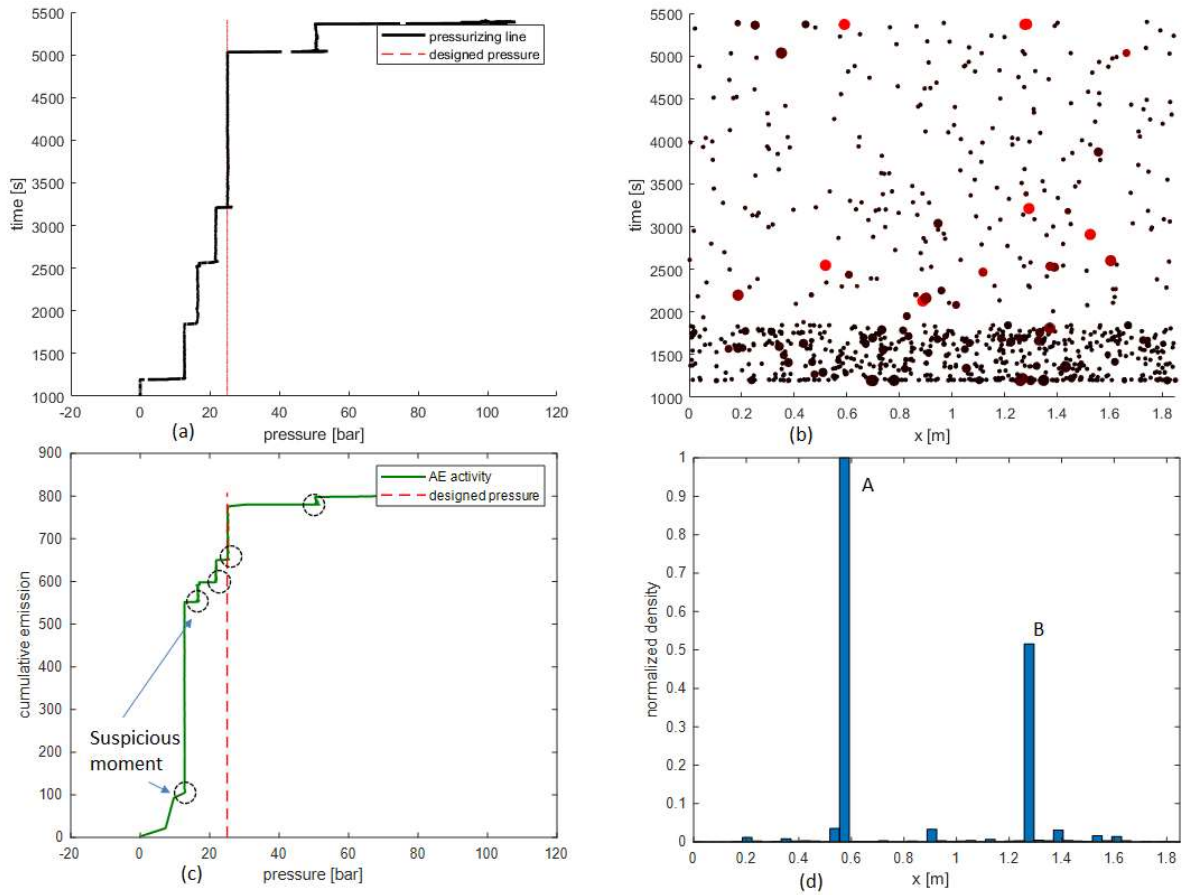
Figure 5. 12. AE source monitoring while pressurizing the testing pipeline: (a) pressurizing process, (b) AE source distribution, (c) AE activity against load, (d) AE source histogram

The source density in a segment is defined as a total number of detected AE sources whose coordinates belong to the segment. Additionally, we determine the source intensity as weights when forming the histogram using the following formula:

$$h_i = h_i + M_s \times N_s \tag{5. 15}$$

where h_i is a current count of sources in the ith segment ($i = 1, 2,\dots , n$), and Ms and Ns are magnitude and duration of a source, respectively. The term $M_s \times N_s$ in Eq. (15) is used to improve the simple counting of sources in a segment, which is $h_i = h_i+1$, thus resulting in a weighted source histogram. With the assistance of the histogram, we can predict irregularities

in the pipeline, at locations where high source density occurs. For example, the A, B points (see Figure 5.12 (d)) are suspicious locations of potential cracks in the histogram, because the emission density is highly concentrated there.

5.4. Experimental Results

Prior to evaluating the algorithm of crack detection and localization, we determine the Rayleigh wave velocity, based on testing signals received by stimulating pencil lead breaking on the surface of the pipeline. Although wave velocity has been specified in Ref. [20, 88], it is necessary to determine this value for a specific object, such as our pipeline, because the manner in which waves propagate through a body or a surface depends on its material, size, and structure. The velocity determination procedure is referred to "wave velocity calibration," as shown in Figure 5.9 in subsection 5.3.2. This is carried out by the following expression:

$$C = \frac{L - 2x}{\Delta t} \qquad (5.\,16)$$

where C is the Rayleigh wave velocity, x is the pencil lead breaking location, L is the distance between the two sensors (1 and 2), and Δt is the TDOA. In fact, Eq. (5. 16) is inversely transformed from Eq. (5. 8).

From multiple pencil lead breaking stimulations at various positions along the pipeline, we calculated the mean value of velocities and found that the applied average velocity was approximately 1873 m/s. Moreover, the paper chooses 79 frequency bands for converting 50000–sample signals from the time domain into the time–frequency domain, $P_{FA} = 0.01\%$ and $N_m = 5$ for the event filtering, $\Delta T_{min} = 50$ and $\Delta F_{min} = 2$ for the event grouping, $e_r = 5\%$ and $n = 50$ for the event source monitoring to obtain the acceptable results.

With the AE signal data recorded in subsection 5.2 during the pressurization process (from the beginning moment until the pipeline was broken) and the method given in subsection 5.3, we obtain the AE source distribution, AE activity against load, and AE source histogram, as depicted in Figure 5.13. It can be seen that major AE sources (large red points in Figure 5.13 (b)) appear when the pressure line increases (Figure 5.13 (a)). This can be explained by the Kaiser effect (the AE activity continues if the current stress rises above the applied former value). In addition, we can clearly see the dependence of AE activity on load in Figure 5.13 (c), and the AE density along the pipeline in Figure 5.13 (d). We note that the density is calculated before the moment the pipeline breaks because the AE sources after the rupture are from the leak, not from a material crack in the pipeline.
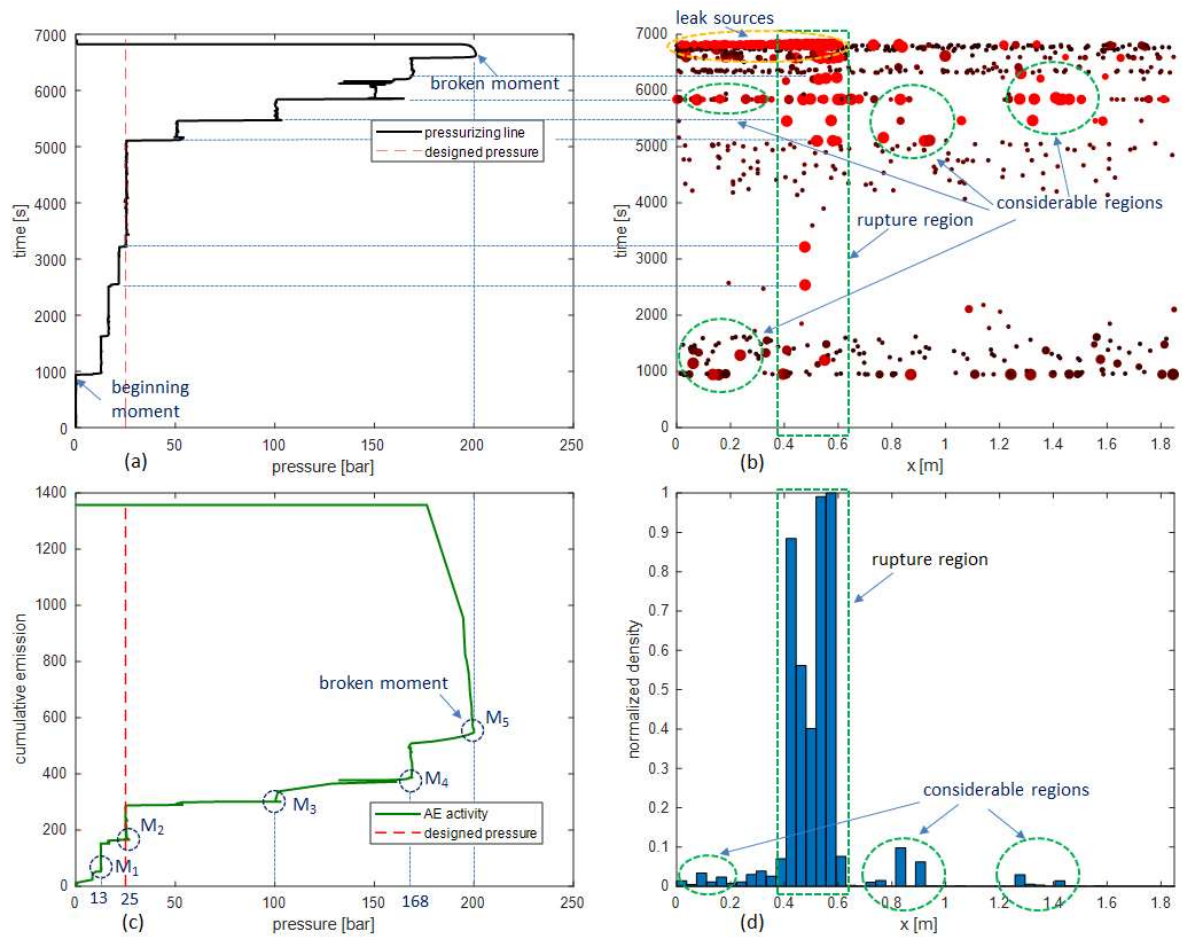
Figure 5. 13. AE source monitoring while pressurizing the testing pipeline until it was broken: (a) pressurizing process, (b) AE source distribution, (c) AE activity against load, (d) AE source histogram
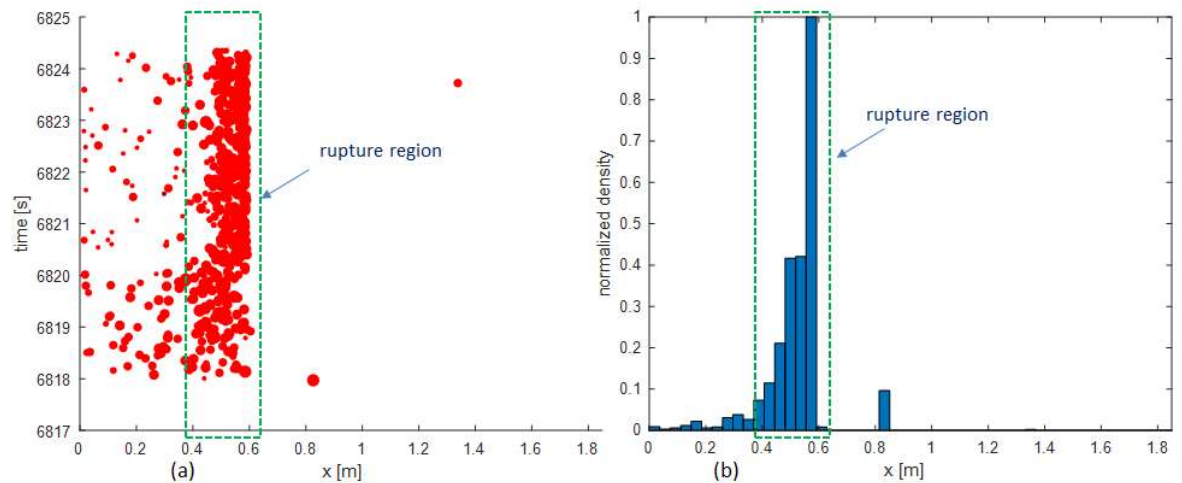
Figure 5. 14. Acoustic emission around M_5: (a) AE source distribution, (b) AE source histogram
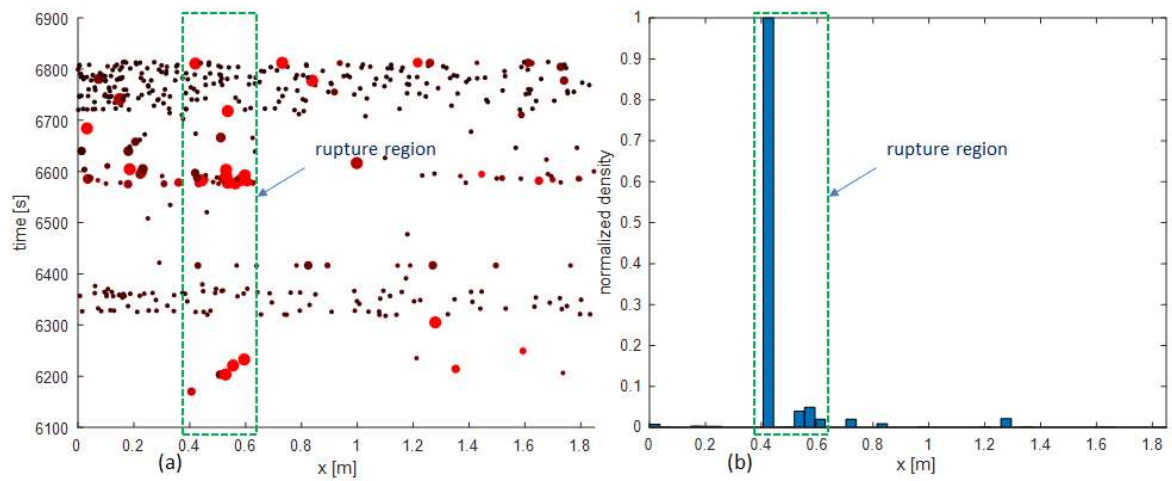
Figure 5. 15. Acoustic emission around M_4: (a) AE source distribution, (b) AE source histogram
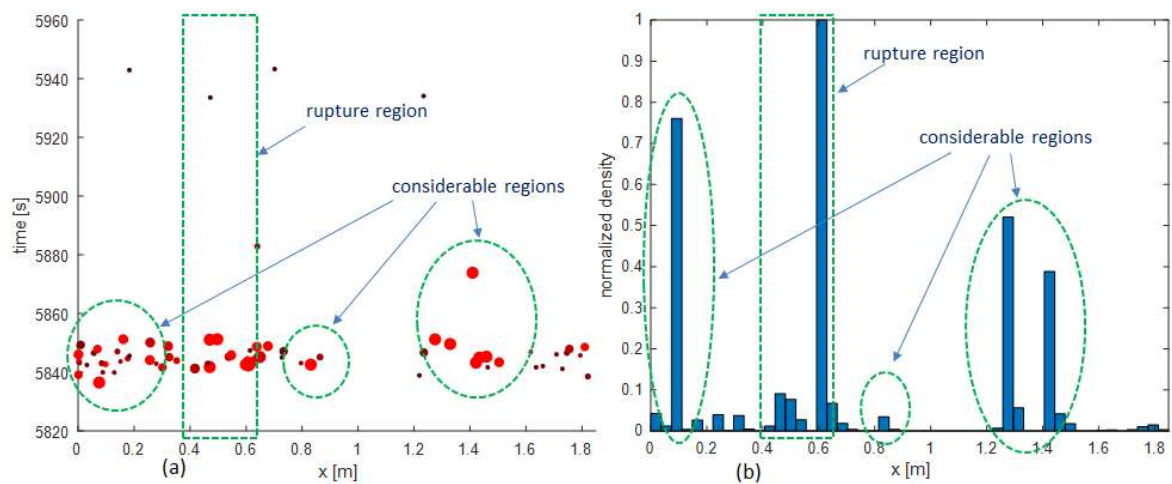
Figure 5. 16. Acoustic emission around M_3: (a) AE source distribution, (b) AE source histogram
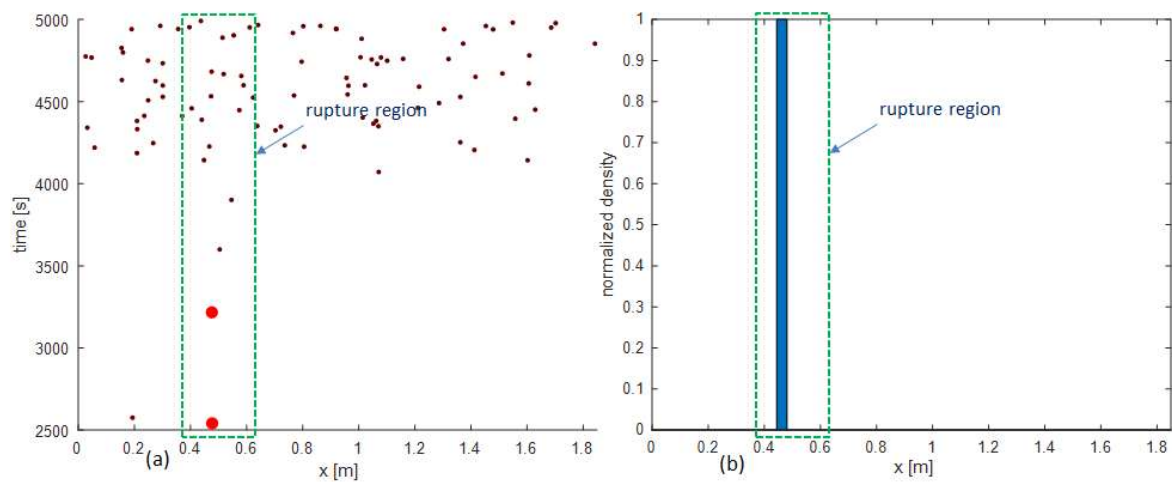
Figure 5. 17. Acoustic emission around M_2: (a) AE source distribution, (b) AE source histogram
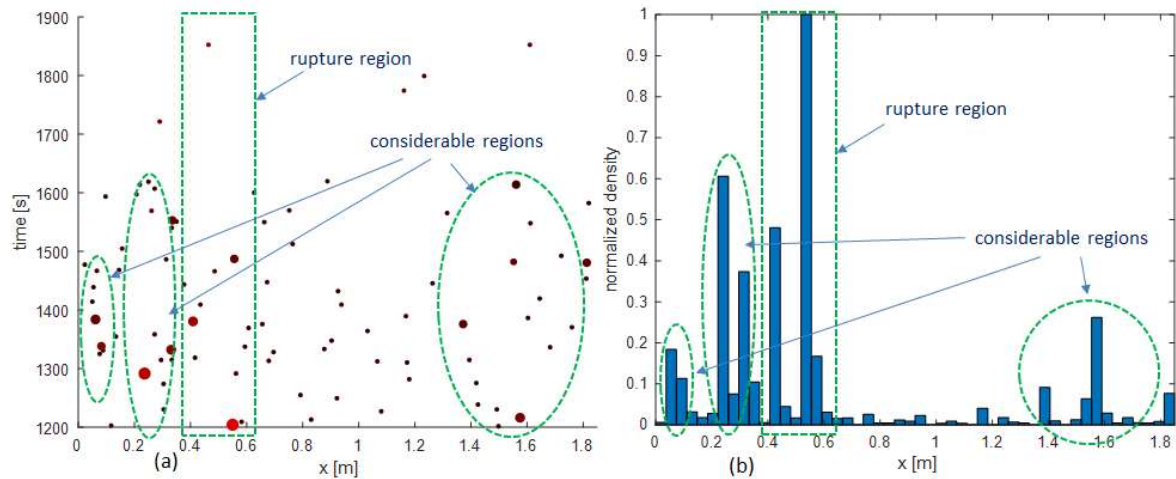
Figure 5. 18. Acoustic emission around M$_1$: (a) AE source distribution, (b) AE source histogram

We conducted the experiment until the pipeline was broken, thus exposing the cracked area. We can therefore mark the rupture region, with a green dashed rectangle, at a fixed coordinate in Figure 5.13 (b and d). Obviously, this is the location of the highest density of sources along the pipeline, as shown in Figure 5.13 (d). In this plot, other positions which are local density peaks are defined as "considerable regions." They should be noted, because they may be imminent ruptures that have not been previously detected. Although not as serious as the known rupture that destroyed the pipeline, other cracks could be growing at considerable regions.

As stated in Subsection 5.3.3, the points M$_1$, M$_2$, M$_3$, M$_4$, and M$_5$ in Figure 5.13 (c) mark irregular structural changes of the pipeline because the AE activity surges while the applied pressure is held. We will subsequently investigate each of these to analyze the crack growth. The M$_5$ position is the rupture moment that we do not want to occur in real situations, and AE sources at this time have not been considered in the histogram in Figure 5.13 (d) because they are created by leakage from the rupture; however, this point is useful for verifying the proposed method of AE source localization. The AE source histogram in Figure 5.14 (b) shows that the rupture region exhibits the highest density of AE sources after the M$_5$ moment. Although there exist plenty of additional AE sources far from the rupture region in Figure 5.14 (a), the AE source density in the rupture region is dominant. The observations of acoustic emission around the M$_1$, M$_2$, M$_3$, and M$_4$ moments in Figure 5.15, Figure 5.16, Figure 5.17, and Figure 5.18 also reveal the concentration of AE sources at the rupture region.

In practice, we would never exploit a pressure exceeding the endurance limitation of pipelines. Consequently, the M$_1$ and M$_2$ moments are more important than the others for early identification of pipeline degradation. With the assistance of variation in AE activity according to pressure, we can detect these moments. Additionally, suspicious locations exhibiting high

densities of AE sources in the AE source histogram indicate where the pipelines should be examined thoroughly. Thus, the proposed method is applicable for crack detection and localization in a fluid pipeline. To emphasize its effectiveness, we implemented a current approach for comparison using a threshold to detect AE hits in the time domain signals [34] and combined this with the Akaike information criterion [72-74] to improve the location accuracy. The implementation is tested using the same signals as those that were analyzed in our method. That technique creates AE hits in pairs from the two signal channels and localizes AE sources via Eq. 8; however, it does not involve the wave mode determination. The wave mode which AE hits fall in is unknown, thus the applied wave velocity should be a typical value of roughly 3000 m/s given by Ref. [43]. We also establish the distribution and histogram of AE sources obtained by this method according to their coordinates, as depicted in Figure 5.19 and Figure 5.20. The AE sources are indicated by filled circles with the same scaling and color rule as employed in the proposed method.
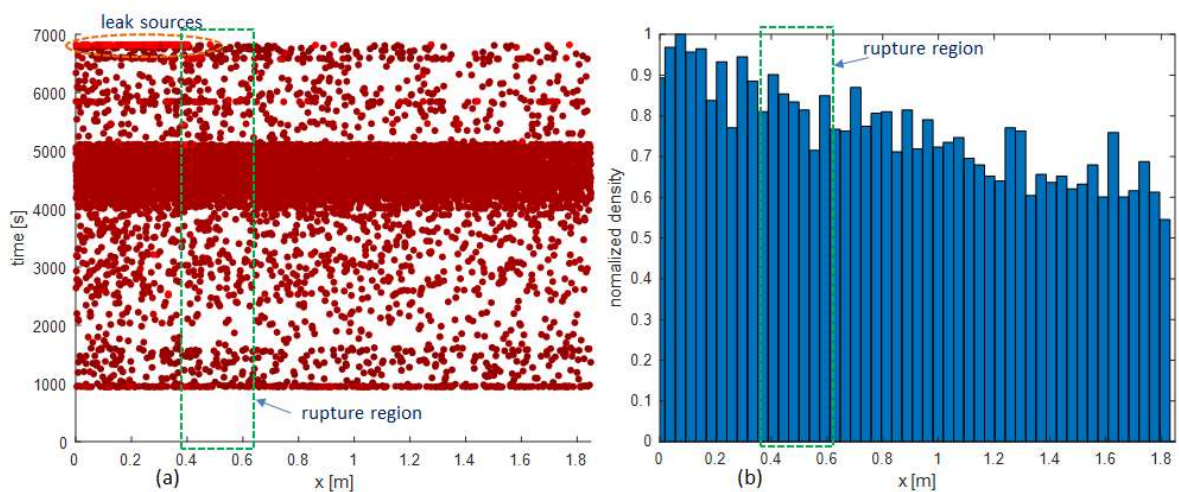
Figure 5. 19. Acoustic emission from the beginning moment of pressurization to the broken moment using a threshold technique to detect AE hits in time domain signals: (a) AE source distribution (b) AE source histogram
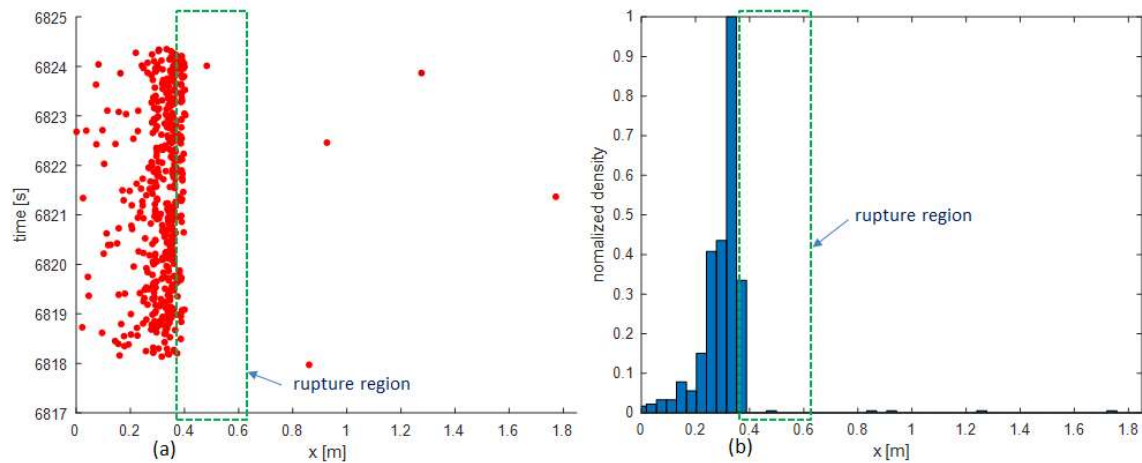
Figure 5. 20. Acoustic emission after the broken moment using a threshold technique to detect AE hits in time domain signals: (a) AE source distribution, (b) AE source histogram

In Figure 5.19 (a), it is challenging to spot the location of the "considerable regions," even though the rupture region is already known. Besides, Figure 5.19 (b) shows the trend of AE source occurrence to the left of the pipeline (whose coordinate is close to zero) while the real fracture is not present. Furthermore, we can observe the most explicit AE sources (Figure 5.19 (a)), which are from leaks due to the rupture. Although these AE sources are localized around a certain coordinate, it does not, however, match with the rupture region as shown in Figure 5.20 (a and b). Hence, the effectiveness of the proposed method is verified because it results in the correct AE source concentration with the correct coordinates (i.e., the fracture location).

5.5. Conclusions

The proposed method showed potential for crack detection and localization in a fluid pipeline. This approach investigated AE signals in the time–frequency domain to locate events. These events were carefully filtered according to the amplitude, frequency, and wave mode. The result was accurately determining the location of emission sources. Based on event sources that are precisely determined in terms of the coordinates and time of the occurrence, the source distribution, the AE activity against load, and the source histogram were constructed to inspect pipeline health. Those can be used to determine not only when a serious structural deformation occurs, but also irregular locations that need to be early checked. The experimental results confirmed the effectiveness of this approach using the data record of an experiment in which we pressurized a water pipeline from its intact status until it ruptured. The proposed method outperformed a conventional one, while exploiting the same datasets. The conventional method only supports an offline signal analysis, whereas the proposed method can be applied online, for a working pipeline, with some adjustments in data acquisition and signal processing.

Chapter 6

Summary of Contributions and Future Work

6.1. Summary of Contributions

The dissertation has focused on pipeline leak and crack identification using AE signals to achieve the objectives: (i) improving reliability and robustness of leak detection, (ii) increasing accuracy of leak localization, (iii) crack detection and localization. The contributions of this thesis are briefly given below:

A scheme of water pipeline leak detection was introduced in Chapter 2. It includes an intermediate step in which AE signals are preprocessed prior to the feature extraction. This step applies the equation of wave attenuation to the signals to compute the function $g(r)$. The use of quantity $g(r)$ is more robust than using direct AE signals in leak detection. The experimental results indicates that the $g(r)$–based method yields accuracy and reliability higher than which employs direct AE signals. This framework is applicable for diverse systems because the behavior of $g(r)$ reflecting a leak is the same. This Chapter also offered a way to choose an appropriate size of window for the signal segmentation. The mechanism determines a possible lag time between the signals channels exploiting the propagation velocity of waves.

Chapter 3 devised a method of leak localization for an industrial–fluid pipeline using AE bursts. This approach includes several stages. First, AE bursts are detected in individual signal channels through thresholds which are adapted to measurements. To further refine the thresholds, the removal of peaks in the signals is suggested before estimating their true level of background–noise. Moreover, adjacent bursts which might be separated by wave dispersion are grouped to decrease the false alarm rate of burst detection. Next, event sources of bursts from two signal channels are localized by the time–delay technique in which the arrival time of bursts is determined at the points on the signals that the thresholds cross. In addition to enhance the localization, a technique of event filtering is also offered. Finally, any anomalies that can be leaks are pinpointed by their high density on a histogram where the event source distribution is illustrated in coordinate along the testing pipeline. The proposed leak localization methodology exceeds the previous ones which employ cross–correlation or generalized cross–correlation functions to estimate the time difference of arrival despite the integration between them and advanced signal denoising techniques such as Wavelet and EMD.

In Chapter 4, a real–time system is developed to detect a leak in a gas pipeline. This system includes a leak detection methodology relying on a k–NN classifier and executes on an MCU–

based platform. The classifier seeks in the hybrid feature pool for the signature closest to inputs using the Manhattan distance. To improve its robustness, the feature pool is normalized utilizing the statistical arguments (mean and standard deviation) of the observations corresponding to the normal state, and then it is purified via the three–sigma rule and the KL distance. The Chapter also offers employing ALEOR to judge the pipeline health to reduce the false alarm rate. In the experiments, the system achieved high accuracy and high robustness of leak detection in the short timeframe; hence, it is applicable for real world applications.

Chapter 5 offered a method for crack detection and localization in a fluid pipeline using AE signals. The approach localizes event sources in the time–frequency domain signals. To enhance the location accuracy, a technique constructed on amplitude, frequency, and wave mode is exploited to filter out false events. Next, the method monitors pipeline health through the source distribution, the AE activity against load, and the source histogram which are constructed on the coordinates and time of the occurrence of event sources. The proposed approach was verified using the recorded data, which outperforms a conventional one. The potential of this method can also be confirmed by its online capability for a working pipeline, in which the data acquisition and the signal–processing program have some modifications.

6.2. Future work

The dissertation has offered potential AE signal–based methods for pipeline leak and crack identification. However, their evaluation only happened in the laboratory using the rather short pipelines. A real pipeline network is usually composed of long pipes. As stated in the previous sections, the signal distortion would become serious on such pipelines because AE waves are subjected to dispersion and attenuation. Moreover, the interference of ambient noise sources, which are abundant near working pipelines, can strongly impact on the signals. Accordingly, the AE portrait resulting from a pipeline in a practical environment may significantly differ from which in the laboratory. Thus, a larger pipeline system will be designed to experiment the methods proposed in this thesis in order to make necessary modifications and achieve those are more applicable for real situations. Because an experiment of crack growth requires more complicated setup phases, especially, in an industrial environment, the future study will primarily focus on the leak identification for pipelines.

On the other hand, the prerequisite of the methods presented in Chapter 2 and Chapter 4 is a priori information about the normal state of pipelines. This is sometimes impossible for a real pipeline system which has been working for ages. Hence, the techniques introduced in Chapter 3 and Chapter 5 can be more appropriate for such cases, which do not require any prerequisite

knowledge of pipeline health. However, their kernels that are the AE event detection and filtering need to be further optimized because a true event related to any anomaly may deeply be concealed in the signals acquired in a real pipeline. In future work, other advanced mechanisms of signal processing and pattern recognition will be adopted to transcend their limitation towards a complete method of leak and crack identification using AE signals, which can be implemented in real pipeline networks.

In addition, to facilitate the study of pipeline leak detection, the digital twin technology can be applied. This technology virtually represents a real–time digital counterpart of a physical object or process. Thus, based on it, both the data collection and the test of leak detection algorithm can be conducted in real time.

Publications

International Journal Papers

1. **Thang Bui Quy**, Sohaib Muhammad, and Jong–Myon Kim. "A Reliable Acoustic EMISSION Based Technique for the Detection of a Small Leak in a Pipeline System." *Energies*. **2019**, 12, 1472. https://doi.org/10.3390/en12081472.

2. **Thang Bui Quy** and Jong–Myon Kim. "Leak localization in industrial-fluid pipelines based on acoustic emission burst monitoring." *Measurement*. **2020**, 151, p.107150.

3. **Thang Bui Quy** and Jong–Myon Kim. "Leak detection in a gas pipeline using spectral portrait of acoustic emission signals." *Measurement*. **2020**, 152, p.107403.

4. **Thang Bui Quy** and Jong–Myon Kim. "Real–time Leak Detection for a Gas Pipeline using a *k*–NN Classifier and Hybrid AE Features." *Sensors*. **2021**, 21, 367. https://doi.org/10.3390/s21020367.

5. **Thang Bui Quy** and Jong–Myon Kim. "Crack detection and localization in a fluid pipeline based on acoustic emission signals." *Mechanical Systems and Signal Processing*. **2021**, 150, p.107254.

Book Chapters

1. **Thang Bui Quy** and Jong–Myon Kim. "Leakage Detection of Water–Induced Pipelines Using Hybrid Features and Support Vector Machines." *Advances in Computer Communication and Computational Sciences, Springer*, **2019**, pp. 377-387.

International Conferences

1. **Thang Bui Quy** and Jong–Myon Kim. "Leakage Detection of Water–Induced Pipelines Using Hybrid Features and Support Vector Machines." *3rd International Conference on Computer, Communication and Computational Sciences*, Bangkok, Thailand, **October 2018**.

Domestic Conferences

1. **Thang Bui Quy** and Jong–Myon Kim. "Leak scale estimation through classification based on vibration." *The 2018 Korean Engineering – Safety – Health – Art Society Conference in Korea (KESHA 2018)*, Ulsan, South Korea, **2018**.

References

[1] JWRC, " Technical report on enhancement of pipe renewal (e-pipe) project for sustainable water service," 2011. [Online]. Available: http://www.jwrc-net.or.jp/shuppan/hakkou-tosho.html

[2] E. W. McAllister, *SEVENTHEDITION PIPELINE RULES OF THUMB HANDBOOK Quick and accurate solutions to your everyday pipeline problems*. Elsevier Inc. All rights reserved., 2009.

[3] G. A. Antaki, *Piping and Pipeline Engineering Design, Construction, Maintenance, Integrity, and Repair*. USA: Marcel Dekker, Inc., 2003.

[4] EGIG, "Gas pipeline incidents," 2020.

[5] P. S. Murvay and I. Silea, "A survey on gas leak detection and localization techniques," *Journal of Loss Prevention in the Process Industries,* vol. 25, no. 6, pp. 966-973, 2012.

[6] S. Gholizadeh, Z. Leman, and B. H. T. Baharudin, "A review of the application of acoustic emission technique in engineering," *Structural Engineering and Mechanics,* vol. 54, no. 6, pp. 1075-1095, 2015.

[7] S. Datta and S. Sarkar, "A review on different pipeline fault detection methods," *Journal of Loss Prevention in the Process Industries,* vol. 41, pp. 97-106, 2016.

[8] C. S. C. a. X. Z. T. K. Chan, "Review of Current Technologies and Proposed Intelligent Methodologies for Water Distributed Network Leakage Detection," *IEEE Access,* vol. 6, pp. 78846-78867, 2018, doi: 10.1109/ACCESS.2018.2885444.

[9] M. S. A. Moubayed, M. Luccini, S. Primak and A. Shami, "Water Leak Detection Survey: Challenges & Research Opportunities Using Data Fusion & Federated Learning," *IEEE Access,* vol. 9, pp. 40595-40611, 2021, doi: 10.1109/ACCESS.2021.3064445.

[10] U. Baroudi, A.-R. A.A., and A. Devendiran, "Pipeline leak detection systems and data fusion: A survey," *IEEE Access,* vol. 7, pp. 97426-97439, 2019.

[11] R. K. Miller, E. v. K. Hill, and P. O. Moore, *Acoustic Emission Testing*. Columbus, OH: American Society for Nondestructive Testing, 2005.

[12] A. Soldevila, Fernandez-Canti, R.M., Blesa, J., Tornil-Sin, S. and Puig, V., "Leak localization in water distribution networks using Bayesian classifiers," *Journal of Process Control,* vol. 55, pp. 1-9, 2017.

[13] Y. Wu, Liu, S., Smith, K. and Wang, X., " Using correlation between data from multiple monitoring sensors to detect bursts in water distribution systems," *Journal of Water Resources Planning and Management,* vol. 144, no. 2, p. 04017084, 2018.

[14] S. R. Mounce, Mounce, R.B. and Boxall, J.B., "Novelty detection for time series data analysis in water distribution systems using support vector machines," *Journal of hydroinformatics,* vol. 13, no. 4, pp. 672-686, 2011.

[15] C. Xu, S. Du, P. Gong, Z. Li, G. Chen, and G. Song, "An Improved Method for Pipeline Leakage Localization With a Single Sensor Based on Modal Acoustic Emission and Empirical Mode Decomposition With Hilbert Transform," *IEEE Sensors Journal,* vol. 20, no. 10, pp. 5480-5491, 2020-05-15 2020, doi: 10.1109/jsen.2020.2971854.

[16] N. K. Banjara, S. Sasmal, and S. Voggu, "Machine learning supported acoustic emission technique for leakage detection in pipelines," *International Journal of Pressure Vessels and Piping,* 2020.

[17] R. Xiao, Q. Hu, and J. Li, "Leak detection of gas pipelines using acoustic signals based on wavelet transform and Support Vector Machine," *Measurement,* vol. 146, pp. 479-489, 2019-11-01 2019, doi: 10.1016/j.measurement.2019.06.050.

[18] A. K. Das, Suthar, D. and Leung, C.K., "Machine learning based crack mode classification from unlabeled acoustic emission waveform features," *Cement and Concrete Research,* vol. 121, pp. 42-57, 2019.

[19] Y. Yan, Y. Shen, X. Cui, and Y. Hu, "Localization of Multiple Leak Sources Using Acoustic Emission Sensors Based on MUSIC Algorithm and Wavelet Packet Analysis," *IEEE Sensors Journal,* vol. 18, no. 23, pp. 9812-9820, 2018-12-01 2018, doi: 10.1109/jsen.2018.2871720.

[20] M. A. Meyers, *Dynamic Behavior of Materials*. New York: Wiley, 1994.

[21] S. W. Rienstra and A. Hirschberg, *An Introduction to Acoustics*. Eindhoven University of Technology, 2018.

[22] J. Sun, Q. Xiao, J. Wen, and F. Wang, "Natural gas pipeline small leakage feature extraction and recognition based on LMD envelope spectrum entropy and SVM," *Measurement,* vol. 55, pp. 434-443, 2014-09-01 2014, doi: 10.1016/j.measurement.2014.05.012.

[23] J. Sun, Q. Xiao, J. Wen, and Y. Zhang, "Natural gas pipeline leak aperture identification and location based on local mean decomposition analysis," *Measurement,* vol. 79, pp. 147-157, 2016-02-01 2016, doi: 10.1016/j.measurement.2015.10.015.

[24] C. Guo, Y. Wen, P. Li, and J. Wen, "Adaptive noise cancellation based on EMD in water-supply pipeline leak detection," *Measurement,* vol. 79, pp. 188-197, 2016-02-01 2016, doi: 10.1016/j.measurement.2015.09.048.

[25] Y. Gao, Brennan, M.J., Y. Liu, F. C. Almeida, and P. F. Joseph, "Improving the shape of the cross-correlation function for leak detection in a plastic water distribution pipe using acoustic signals," *Applied Acoustics,* vol. 127, pp. 24-33, 2017.

[26] Q. Xiao, J. Li, J. Sun, H. Feng, and S. Jin, "Natural-gas pipeline leak location using variational mode decomposition analysis and cross-time–frequency spectrum," *Measurement,* vol. 124, pp. 163-172, 2018-08-01 2018, doi: 10.1016/j.measurement.2018.04.030.

[27] Z. Li, Z. H., D. Tan, X. Chen, and H. Lei, "A novel acoustic emission detection module for leakage recognition in a gas pipeline valve," *Process Safety and Environmental Protection,* vol. 105, pp. 32-40, 2017.

[28] L. Yu and S. Z. Li, "Acoustic emission (AE) based small leak detection of galvanized steel pipe due to loosening of screw thread connection," *Applied Acoustics,* vol. 120, pp. 85-89, 2017.

[29] S. Li, Y. Song, and G. Zhou, "Leak detection of water distribution pipeline subject to failure of socket joint based on acoustic emission and pattern recognition," *Measurement,* vol. 115, pp. 39-44, 2018-02-01 2018, doi: 10.1016/j.measurement.2017.10.021.

[30] M. Nicola, Nicola, C.I., Vintila, A., Hurezeanu, I. and Duta, M., "Pipeline Leakage Detection by Means of Acoustic Emission Technique Using Cross-Correlation Function," *J. Mech. Eng. Autom,* vol. 8, pp. 59-67, 2018.

[31] Y. J. Song and S. Z. Li, "Leak detection for galvanized steel pipes due to loosening of screw thread connections based on acoustic emission and neural networks," *Journal of Vibration and Control,* vol. 24, no. 18, pp. 4122-9, 2018.

[32]   X. Diao *et al.*, "An improved variational mode decomposition method based on particle swarm optimization for leak detection of liquid pipelines," *Mechanical Systems and Signal Processing,* vol. 143, p. 106787, 2020-09-01 2020, doi: 10.1016/j.ymssp.2020.106787.

[33]   Q. Ai, Liu, C.X., Chen, X.R., He, P. and Wang, Y., "Acoustic emission of fatigue crack in pressure pipe under cyclic pressure," *Nuclear Engineering and Design,* vol. 240, no. 10, pp. 3616-3620, 2010.

[34]   H. Vallen, *AE testing Fundamentals, Equipment, Applications*. Icking (Munich), Germany: The Acoustic Emission Company, 2002.

[35]   S. K. Mandal, F. T. S. Chan, and M. K. Tiwari, "Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM," *Expert Systems with Applications,* vol. 39, no. 3, pp. 3071-3080, 2012-02-01 2012, doi: 10.1016/j.eswa.2011.08.170.

[36]   A. Martini, A. Rivola, and M. Troncossi, "Autocorrelation Analysis of Vibro-Acoustic Signals Measured in a Test Field for Water Leak Detection," *Applied Sciences,* vol. 8, no. 12, p. 2450, 2018. [Online]. Available: https://www.mdpi.com/2076-3417/8/12/2450.

[37]   P. He, "Simulation of ultrasound pulse propagation in lossy media obeying a frequency power law," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control,* vol. 45, no. 1, pp. 114-125, 1998-01-01 1998, doi: 10.1109/58.646916.

[38]   T. L. Szabo and J. Wu, "A model for longitudinal and shear wave propagation in viscoelastic media," *The Journal of the Acoustical Society of America,* vol. 107, no. 5, pp. 2437-2446, 2000.

[39]   W. Chena and S. Holm, "Modified Szabo's wave equation models for lossy media obeying frequency power law," *The Journal of the Acoustical Society of America,* vol. 114, no. 5, pp. 2570-2574, 2003.

[40]   S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 2 ed. USA: Academic Press, 2003.

[41]   M. Group, "R15I-AST Sensor," Princeton Junction, NJ, USA, 2015.

[42]   B. E. 15856, *Non-Destructive Testing – Acoustic Emission – General Principles of AE Testing for the 416 Detection of Corrosion with Metallic Surrounding Filled with Fluid*. European Standard, 2010.

[43]   M. Control, "Sound Speed & Pipe Size Data," 2014.

[44]   G. Z. Watters, "Analysis and Control of Unsteady Flow in Pipelines," 1984.

[45]   J. Twyman, "Wave speed calculation for water hammer analysis," *Obras y Proyectos,* 2016.

[46]   L. Devroye, L. Györfi, and G. Lugosi, *A probabilistic theory of pattern recognition*. Springer Science & Business Media, 2013.

[47]   B. Muravin, "Acoustic emission science and technology," *The journal of Building and infrastructure engineering of the Israeli association of engineers and architects,* vol. 1, pp. 4-5, 2009.

[48]   T. L. Szabo, "Time domain wave equations for lossy media obeying a frequency power law," *The Journal of the Acoustical Society of America,* vol. 96, no. 1, pp. 491-500, 1994.

[49] Y. Gao, Brennan, M.J., Joseph, P.F., Muggleton, J.M. and Hunaidi, O., "A model of the correlation function of leak noise in buried plastic pipes," *Journal of Sound and Vibration,* vol. 277, no. 1-2, pp. 133-148, 2004.

[50] L. Chen, Liu, Y., Kong, F. and He, N., "Acoustic source localization based on generalized cross-correlation time-delay estimation," *Procedia engineering,* vol. 15, pp. 4912-4919, 2011.

[51] Y. Gao, Brennan, M.J., Liu, Y., Almeida, F.C. and Joseph, P.F., "Improving the shape of the cross-correlation function for leak detection in a plastic water distribution pipe using acoustic signals," *Applied Acoustics,* vol. 127, pp. 24-33, 2017.

[52] C. a. C. Knapp, G., "The generalized correlation method for estimation of time delay," *IEEE transactions on acoustics, speech, and signal processing,* vol. 24, no. 4, pp. 320-327, 1976.

[53] S. Pan, Xu, Z., Li, D. and Lu, D., "Research on detection and location of fluid-filled pipeline leakage based on acoustic emission technology," *Sensors,* vol. 18, no. 11, p. 3628, 2018.

[54] C. Guo, Wen, Y., Li, P. and Wen, J., "Adaptive noise cancellation based on EMD in water-supply pipeline leak detection," *Measurement,* vol. 79, pp. 188-197, 2016.

[55] L. B. a. B. White, B., "Cross spectral analysis of nonstationary processes," *IEEE Transactions on Information Theory,* vol. 36, no. 4, pp. 830-835, 1990.

[56] J. S. Bendat and A. G. Piersol, *Random Data Analysis and Measurement Procedures.* Hoboken, New Jersey: Wiley, 2010.

[57] S. M.Kay, *Detection Theory* (Fundamentals of Statistical Signal Processing). Upper Saddle River, New Jersey 07458: Prentice Hall PTR, 1998.

[58] S. V. Ranganayakulu, B. S. Goud, P. V. Sastry, and B. R. Kumar, "Calibration of acoustic emission system for materials characterization," *Univers. J. Mater. Sci,* vol. 3, pp. 62-69, 2015.

[59] L. Wang, X. Gao, and T. Liu, "Gas pipeline small leakage feature extraction based on LMD envelope spectrum entropy and PCA–RWSVM," *Transactions of the Institute of Measurement and Control,* vol. 38, no. 12, pp. 1460-1470, 2016.

[60] L. Q. Dong, Z.; Wang, H.; Yang, W.; Zhao, W.; Xu, K.; Wang, G.; Zhao, L.; Yan, H., "The gas leak detection based on a wireless monitoring system," *IEEE Trans. Ind. Inf.,* vol. 15, pp. 6240–6251, 2019.

[61] F. B. E. Chraim, Y.; Pister, K., "Wireless gas leak detection and localization," *IEEE Trans. Ind. Inf.,* vol. 12, pp. 768–779, 2016.

[62] F. Pukelsheim, *The three sigma rule.* The American Statistician, 1994.

[63] A. V. Oppenheim and R. W. Schafer, *Discrete-time signal processing*, 3 ed. Pearson, 2010.

[64] D. S. Lemons, *An Introduction to Stochastic Processes in Physics.* USA: The Johns Hopkins University Press, 2002.

[65] G. a. O. Ch, M., "Acoustic Emission Testing," in *Basic for Research-Applications in Civil Engineering*, Berlin Heidelberg, 2008: Springer Verlag.

[66] *Practical Acoustic Emission Testing.* Japan: The Japanese Society for Non-Destructive Inspection, 2006.

[67] D. Karaduman, Bircan, D.A. and Çetin, A., "Non-Destructive examination of underground pressure vessels using acoustic emission (AE) techniques," *European Mechanical Science,* vol. 1, no. 1, pp. 1-8, 2017.

[68] T. a. T. Roberts, M., "Acoustic emission monitoring of fatigue crack propagation," *Journal of constructional steel research,* vol. 59, no. 6, pp. 695-712, 2003.

[69] A. Behnia, Ranjbar, N., Chai, H.K. and Masaeli, M., "Failure prediction and reliability analysis of ferrocement composite structures by incorporating machine learning into acoustic emission monitoring technique," *Construction and Building Materials,* vol. 122, pp. 823-832, 2016.

[70] H. Vallen, "AE testing Fundamentals, Equipment, Applications," 2002.

[71] T. L. a. W. Szabo, J., "A model for longitudinal and shear wave propagation in viscoelastic media," *The Journal of the Acoustical Society of America,* vol. 107, no. 5, pp. 2437-2446, 2000.

[72] Z. Zhou, Cheng, R., Rui, Y., Zhou, J. and Wang, H., "An improved automatic picking method for arrival time of acoustic emission signals," *IEEE Access,* vol. 7, pp. 75568-75576, 2019.

[73] F. Sagasta, Tee, K.F. and Piotrkowski, R., "Lamb modes detection using cumulative Shannon Entropy with improved estimation of arrival time," *Journal of Nondestructive Evaluation,* vol. 38, no. 1, pp. 1-16, 2019.

[74] F. P. Moctezuma, Prieto, M.D. and Martínez, L.R., "Performance analysis of acoustic emission hit detection methods using time features," *IEEE access,* vol. 7, pp. 71119-71130, 2019.

[75] M. Barroso-Romero, Gagar, D., Pant, S. and Martinez, M., "Wave mode identification of acoustic emission signals using phase analysis," *Acoustics,* vol. 1, no. 2, pp. 450-472, 2019.

[76] M. R. Pearson, Eaton, M., Featherston, C., Pullin, R. and Holford, K., "Improved acoustic emission source location during fatigue and impact events in metallic and composite structures," *Structural Health Monitoring,* vol. 16, no. 4, pp. 382-399, 2017.

[77] F. Bai, Gagar, D., Foote, P. and Zhao, Y., "Comparison of alternatives to amplitude thresholding for onset detection of acoustic emission signals," *Mechanical Systems and Signal Processing,* vol. 84, pp. 717-730, 2017.

[78] E. Kalkan, "An automatic P-phase arrival-time picker," *Bulletin of the Seismological Society of America,* vol. 106, no. 3, pp. 971-986, 2016.

[79] S. K. Al-Jumaili, Pearson, M.R., Holford, K.M., Eaton, M.J. and Pullin, R., "Acoustic emission source location in complex structures using full automatic delta T mapping technique," *Mechanical Systems and Signal Processing,* vol. 72, pp. 513-524, 2016.

[80] M. A. a. F. Torres-Arredondo, C.P., "Characterization and classification of modes in acoustic emission based on dispersion features and energy distribution analysis," *Shock and Vibration,* vol. 19, no. 5, pp. 825-833, 2012.

[81] M. G. Baxter, Pullin, R., Holford, K.M. and Evans, S.L., "Delta T source location for acoustic emission," *Mechanical systems and signal processing,* vol. 21, no. 3, pp. 1512-1520, 2007.

[82] M. a. A. Ettouney, S., *Infrastructure health in civil engineering.* London. UK: CRC press, 2012.

[83] M. Group, "Express-8 Product Bulletin," 2013.

[84] R. J. E. a. S. Merry, M., *Wavelet theory and applications*. Eindhoven university of technology, Department of mechanical engineering, Control systems technology group, 2005.

[85] J. a. L. Feng, H., "Peak analysis of grayscale image: algorithm and application," *International Journal of Information Technology,* vol. 12, no. 5, pp. 11-8, 2006.

[86] L. J. S., *Two-Dimensional Signal and Image Processing*. Englewood Cliffs, NJ: Prentice Hall, 1990.

[87] M. A. Hamstad, "SOME OBSERVATIONS ON RAYLEIGH WAVES AND ACOUSTIC EMISSION IN THICK STEEL PLATES," *Journal of Acoustic Emission,* vol. 27, 2009.

[88] L. B. Freund, *Dynamic fracture mechanics*. Cambridge university press, 1998.