December 2022

Multiple Surface Pipeline Leak Detection Using Real-Time Sensor Data Analysis

Francis Enejo Idachaba

MULTIPLE SURFACE PIPELINE LEAK DETECTION USING REAL-TIME SENSOR DATA ANALYSIS

BY

Francis Enejo Idachaba

Bachelor of Engineering, Federal University of Technology Yola, 1997

Master of Engineering, University of Benin, 2002

Doctor of Philosophy, University of Benin 2009

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Grand Forks, North Dakota

December

2022

This thesis, submitted by Francis Enejo Idachaba in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Petroleum Engineering from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Dr. Olusegun Tomomewo

Name of Chairperson

Professor Sven Egenhoff

Name of Committee Member

Dr. Wafik Beydoun

Name of Committee Member

Dr Kegang Ling

Name of Committee Member

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

____

Chris Nelson

Dean of the School of Graduate Studies

____

Date

PERMISSION

Title Multiple Surface Pipeline Leak Detection Using Real-Time Sensor Data Analysis

Department Petroleum Engineering

Degree Doctor of Philosophy

In presenting this thesis in partial fulfilment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and the University of North Dakota for any scholarly use which may be made of any material in my thesis.

Francis Enejo Idachaba

1st December 2022

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to my advisor, Dr. Olusegun Tomomewo for stepping in and providing all the needed guidance and support, I am deeply grateful for your commitment to this work and all the push you provided to help me finish this work. I also want to appreciate my previous adviser Rabiei Minou and chair Professor Vamegh for their guidance and support. I want to appreciate the College of Engineering and Mines and the North Dakota Industrial Commission for the funding provided to enable me to undertake this research. I wish to also appreciate the Dean and the acting chair of the department for all the support they continue to provide to enable me to complete this program. I am deeply grateful. I would like to extend my appreciation to the committee members, Dr. Olusegun Tomomewo, Dr. Kegang Ling, Professor Sven Egenhoff, and Dr. Wafik Beydoun for stepping in at such short notice to enable me to complete this research. I want to use this opportunity to appreciate the support of my wife, for enduring all the work across multiple time zones and the countless hours spent away from home taking the classes and doing the research.

Dedication

To my mom and my dad both of whom passed on before I could finish this program

ABSTRACT

Pipelines enable the largest volume of both intra and international transportation of oil and gas and play critical roles in the energy sufficiency of countries. The biggest drawback with the use of pipelines for oil and gas transportation is the problem of oil spills whenever the pipelines lose containment. The severity of the oil spill on the environment is a function of the volume of the spill and this is a function of the time taken to detect the leak and contain the spill from the pipeline. A single leak on the Enbridge pipeline spilled 3.3 million liters into the Kalamazoo river while a pipeline rupture in North Dakota which went undetected for 143 days spilled 29 million gallons into the environment.

Several leak detection systems (LDS) have been developed with the capacity for rapid detection and localization of pipeline leaks, but the characteristics of these LDS limit their leak detection capability. Machine learning provides an opportunity to develop faster LDS, but it requires access to pipeline leak datasets that are proprietary in nature and not readily available. Current LDS have difficulty in detecting low-volume/low-pressure spills located far away from the inlet and outlet pressure sensors. Some reasons for this include the following, leak induced pressure variation generated by these leaks is dissipated before it gets to the inlet and outlet pressure sensors, another reason is that the LDS are designed for specific minimum detection levels which is a percentage of the flow volume of the pipeline, so when the leak falls below the LDS minimum detection value, the leak will not be detected. Perturbations generated by small volume leaks are often within the threshold values of the pipeline's normal operational envelop as such the LDS disregards these perturbations. These challenges have been responsible for pipeline leaks going on for weeks only to be detected by third-party persons in the vicinity of the leaks.

This research has been able to develop a framework for the generation of pipeline datasets using the PIPESIM software and the RAND function in Python. The topological data of the pipeline right of way, the pipeline network design specification, and the fluid flow properties are the required information for this framework. With this information, leaks can be simulated at any point on the pipeline and the datasets generated. This framework will facilitate the generation of the One-class dataset for the pipeline which can be used for the development of LDS using machine learning.

The research also developed a leak detection topology for detecting low-volume leaks. This topology comprises of the installation of a pressure sensor with remote data transmission capacity at the midpoint of the line. The sensor utilizes the exception-based transmission scheme where it only transmits when the new data differs from the existing data value. This will extend the battery life of the sensor. The installation of the sensor at the midpoint of the line was found to increase the sensitivity of the LDS to leak-induced pressure variations which were traditionally dissipated before getting to the Inlet/outlet sensors. The research also proposed the development of a Leak Detection as a Service (LDaaS) platform where the pressure data from the inlet and the midpoint sensors are collated and subjected to a specially developed leak detection algorithm for the detection of pipeline leaks. This leak detection topology will enable operators to detect low-volume/low-pressure leaks that would have been missed by the existing leak detection system and deploy the oil spill response plans quicker thus reducing the volume of oil spilled into the environment. It will also provide a platform for regulators to monitor the leak alerts as they are generated and enable them to evaluate the oil spill response plans of the operators.

Contents

Abstract vi
Dedication v
Acknowledgements iv
List Of Figures x
List Of Tables xii
Chapter 1 1
Pipeline Leak Detection 1
1.0 Introduction 1
1.1 Pipeline Leak Detection Strategies 5
1.2 Research Gaps 6
1.3 Research Questions 7
1.4 Research Aim and Objectives 7
1.5 Methodology 8
1.6 Significance 8
1.7. Expected Contribution 8
1.8 Thesis Structure 9
1.9 Summary 9
Chapter 2 11
Literature Review 11
2.0 Low Pressure Pipeline Leak Challenge 11
2.1. Pipeline Transport of Oil and Gas Products 16
2.2 Pipeline Leak Lifecycle 17
2.3 Pipeline Leak Detection Methods 19
2.4 Pipeline Modelling 27
2.4.1 Governing Equations 32
2.4.2 Pipeline pressure profiles 33
2.4.3 Correlation of pipeline pressure to time series data 36
2.5 Anomaly Detection in Time series Data 39
2.5.1 Types of Anomalies (Outliers) 42
2.5.2 Anomaly detection strategies 42
2.5.3 Deep Learning-based Anomaly Detection System for Pipeline Leaks 43
2.5.4 Real-Time and Batch Anomaly Detection 50
2.6 Performance Metrics for Anomaly Detection 52

2.7 Unsupervised learning algorithm for pipeline leak detection 57
2.7.1 Data Analytics Applications for Leak Detection (Selected Case studies) 58
2.8 PIPESIM Simulation Software 59
2.9 One-Class Classifiers.. 61
2.9.1 One class classifier Model: 62
Chapter 3 64
Materials and Methods 64
3.0 Materials and Methods.. 64
3.1 Research Question 1 64
3.1.1 Research Question 1: Methodology 64
3.1.2 Experimental Processes for Research Question 1 65
3.2 Research Question 2 72
3.2.1 Research Question 2: Methodology 72
3.2.2 Experimental Processes for Research Question 2 72
Chapter 4 75
Results 75
4.0 Results for Pipeline leak and No leak Data generation 75
4.1 Results for Research Question 1 75
4.2 Results for Research Question 2 84
Chapter 5 91
Discussion 91
5.0 Discussion.. 91
5.1 Pipeline Leak and No leak Time series Dataset Generation 93
5.2 One class classifier and leak dataset generation 94
5.3 Midpoint Sensor Installation on the Pipeline 95
5.4 Leak Detection as a Service (LDaaS). System Overview 96
Chapter 6 99
Conclusions 99
6.0 Research contributions 99
`6.1 Recommendations for further research 100

List of Figures

Figure 1.1 Pipeline Leak lifecycle 5

Figure 2.1 USA pipeline leak locations and the impact. (Richard 2013) 17

Figure 2.2 Leak Detection Methods (Adegboye et al. 2019). 22

Figure 2.3 Internal Leak detection method. (Baroudi et al. 2019) 24

Figure 2.4 Classification of pipeline leak detection systems (Baroudi et al. 2019) 26

Figure 2.5 Pipeline with the leak at distance D1 from the inflow section 29

Figure 2.6 Pipeline network modelling for leak detection 31

Figure 2.7 Machine Learning pipeline. (Akerkar 2019) 31

Figure 2.8 Moody Diagram (Moody and Princeton 1944). 33

Figure 2.9 Pressure profile for single leak condition 34

Figure 2.10 Leak Localization 35

Figure 2.11 Multiple pipeline leak pressure profiles 36

Figure 2.12 Time series data from pipeline sensors 37

Figure 2.13 Taxonomy of outlier detection techniques (Basu and Meckesheimer 2006). 40

Figure 2.14 Classification of Anomaly Detection techniques (Saranya and Chellammal 2020) 44

Figure 2.15 Block diagram for Applications using HTM High-Order Inference (Ahmad and Hawkins 2017) 49

Figure 2.16 Normal distribution curves showing the anomaly threshold points. (Krithikadatta 2014) 53

Figure 2.17 Area Under the curve 57

Figure 3.1 Inlet parameters 65

Figure 3.2 Outlet parameters 66

Figure 3.3 Pipeline Parameters 66

Figure 3.4 Complete experimental setup for 20km pipeline.. 66

Figure 3.5 Inlet parameters 68

Figure 3.6 Outlet parameters... 69

Figure 3.7. Leak valve selection 69

Figure 3.8 Pipeline Parameters 70

Figure 3.9 Complete experimental setup for 20km pipeline with leak point at 10km.... 70

Figure 3.10 Sequential leaks at multiple locations.. 73

Figure 4.1 Pressure profile of the line with no leaks.. 76

Figure 4.2 Time series Dataset for the No leak condition of the pipeline 78

Figure 4.3 Pressure profile of line with One leak at 10km 79

Figure 4.4 Pressure profile of No leak and One leak simulation experiment 79

Figure 4.5 Progressive pressure drop at a pipeline leak location (midpoint of the pipeline).. 81

Figure 4.6 Time series waveform for pipeline leak dataset 84

Figure 4.7 Leak 2km from the inlet 86

Figure 4.8 Leak 4km from the inlet 86

Figure 4.9 Leak at the midpoint (10km) 87

Figure 4.10 Leak 4km from the outlet 87

Figure 4.11 Leak 2km from the outlet 88

Figure 4.12 Pressure profile for different leak locations 88

Figure 4.13 Sensor sensitivity analysis.. 89

Figure 5.1 Leak Detection as a Service Platform.. 98

List of Tables

Table 1. 1 Characteristics of leak detection strategies (Fiedler 2016) 6

Table 2. 1 Shows some pipeline oil spills and the delays associated with their detection 15

Table 2.2 Pressure profile descriptions 35

Table 2. 3. Comparison of the different Anomaly detection methods (Anodot 2017) ... 39

Table 2.5 Comparison between HTM and Deep Learning Algorithms(Brody 2018) 50

Table 2.6 Anomaly Detection Metrics 55

Table 2.7 A summary of Data Analytic models for leak detection.(Idachaba and Minou 2021) 59

Table 3. 1. Sensor Data 67

Table 4.1 Pipeline pressure profile data 75

Table 4.2 Sensor Data 77

Table 4.3 Leak and No leak pressure profile 80

Table 4. 4 Pipeline Leak pressure profile (Leak at leak point= 740 psi to 350psi) 82

Table 4.5 Pipeline Pressure leak dataset for single leaks at different points 84

Chapter 1

Pipeline Leak Detection

1.0 Introduction

Pipeline leaks are one of the greatest challenges associated with oil and gas production. While the pipelines provide the most economical mode for transporting large volumes of both crude oil and refined products, they present a very high-risk value to the environment whenever there is a leak. These leaks can be either due to failures associated with age, deliberate acts of vandalization or pipe failure due to operational error. (Dey 2004).

The very long distances and varied terrain where petroleum pipelines transverse, complicate the monitoring and maintenance of such pipelines. To overcome these challenges, a variety of techniques have been developed for the monitoring of these pipelines. These methods range from visual inspection (Guo et al. 2009), electromagnetic methods (Wang et al. 2012), acoustic methods (Juliano et al. 2013), ultrasonic, radiographic, and thermographic methods (Zheng and Yehuda 2013), and more recently transient-based inspection (Lee et al. 2008; Gong et al. 2018; Gong et al. 2014).

The proposed methods are typically divided into the exterior, visual/biological methods, and interior/computation. Some authors also classify them into non-optical and optical methods or hardware and software-based methods. Exterior methods include acoustic sensing (Mahmutoglu and Turk 2018), fiber optic sensors (Png et al. 2018), vapor sampling (Boaz et al. 2014), and ground penetration radar (Hoarau et al. 2017).

Visual/biological methods include AUV/Drone, Trained Dog/Human, etc. Interior/computation-based methods include mass-volume balance (Gao et al. 2018), negative pressure waves (Chen et al. 2018), pressure point analysis (He et al. 2017), and dynamic modeling (Xinhong et al. 2018). Several researchers have developed different methods for leak detection with varying characteristics and success levels. Classical references can be found in (Wylie et al. 1993; Henrie et al. 2016) more recently, using inversion methods in (Wu et al. 2015), The acoustic emission method was also reported by Muntakim et al. (2017).

Other methods which comprise the installation of physical sensors such as optic fiber cables along the pipeline provide very reliable and robust leak detection and localization but at a very high cost. The fiber optic sensing, pressure point analysis, and dynamic modeling are capable of simultaneously detecting the leak size and location.

Fluid transient waves provide a robust system for detecting pipeline characteristics during fluid flow (Bohorquez et al. 2018). However, this technique is highly sensitive to multiple system characteristics, and understanding the pressure signal response to different types of faults is often very challenging (Xu and Karney 2017). The transient-based model with very minimal setup requirement and high detection accuracy has been the LDS method of choice as the results can be obtained quickly (Gong et al. 2013; Lee et al. 2006; Shi et al. 2017).

The mass balance method is one of the most popular methods in the industry for the detection and (segment-wise) location of leaks in long pipeline networks. This method utilizes the real-time transient model (RTTM). The transient model is used to determine the expected pressure along the pipeline and also the flow profiles. The actual data acquired from flow sensors installed on the line are then compared with the estimated values and the real-time calculation of the balance rate, defined by the difference of the actual estimated values serves to determine the pipeline mass/volume balance. The difference should be zero when there is no leak and when this difference exceeds a specific threshold, a leak is inferred and the appropriate alarms triggered, indicating the segment in the network where the unbalance is reported. (Whaley et al. 1992).

Some of the challenges associated with the LDS based on the RTTM include the time required for the computation of the pipeline flow parameters using the transient model. This results in delays with the determination of pipeline leaks as the estimators used in the transient model require a long period to carry out the estimation. The estimation error may not converge to zero because these errors propagate along the line and affect the overall predicting performance of the LDS for leak diagnosis.

Other factors which impact on the success of this system include unreliable data acquisition, the unexpected low resolution of pressure sensors and uncontrolled fluctuations are common implementation problems that may lead to false alarms for leak detection systems based on (open loop) transient models (Fukushima et al. 2000; van Reet and Skogman 1987; Modisette 2004; Liou, 1991).

The hardware approach provides the most accurate leak detection and localization system, but its deployment is limited due to the very high costs associated with the sensor devices and the possibility of vandalization. The software-based LDS, on the other hand, relies on leak datasets and Machine learning for the robust Leak detection system with superior performance but this approach relies on the availability of pipeline leak datasets for training the machine learning algorithms. The software-based approach has been shown to provide good performance with high accuracy. The advancement of cloud computing technologies supports the increasing popularity of this method (Adegboye et al. 2019). A major challenge with this approach is the fact that it is difficult to obtain the actual leak data for pipelines.

Dynamic modeling is an alternative that helps to overcome the challenge of access to pipeline data. It is used to create and analyze models that mimic the actual field. The method combines various elements, such as the fluid type, external environment, pipeline material; length; and diameter, and analyzes parameters, such as flow rate, pressure, and temperature through complex relationships.

A hypothetical leak situation is created on the pipeline model and the different parameters such as the flow rate, pressure, and temperature change are analyzed using computer simulations. The time-intensive nature associated with the pipeline network leads to the limited real-time capability of this system. The difference between the actual and the simulated data of the pipeline also introduces some errors into the actual results (He et al. 2017). These errors can be reduced by the introduction of machine learning algorithms that can learn the specific patterns of the pipeline flow using the synthetic data and make leak predictions without the need for explicit programming. (Mitchell 1997; Ian et al. 2016; Koza et al. 1996). This has the capacity of enabling the rapid detection of pipeline leaks when actual pipeline data is used.

The impact of pipeline leaks is amplified by the delays in the detection of these leaks. Some leaks have been known to continue for days and weeks before being detected. The operators' actions also play a part in determining the spread of the leak. The environment of the leak is one other critical parameter that aids the spread of the leak. Leaks in offshore environments spread much faster as the water body serves as a transport medium for the fluid, and this makes it more difficult to contain and more devastating to the environment. Figure 1.1 shows the leak lifecycle of petroleum pipeline leaks

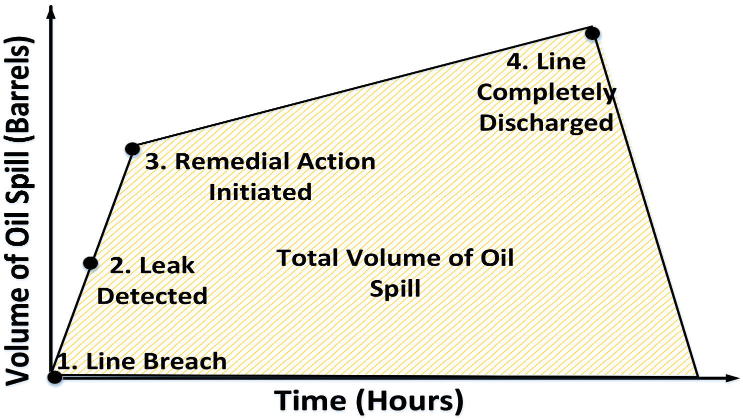


Figure 1.1 Pipeline Leak lifecycle

From the diagram in Figure 1.1, the pipeline discharges its contents into the environment under the pressure in the pipe. This continues well after the leak has been detected until the appropriate remedial action is taken. The remedial actions commence with the shutting down of the line and/or the diversion of the fluid. This leads to a reduction in the discharge rate of the fluid from the leak point and continues till the pipeline section is empty. The volume of crude oil dumped into the environment determines the cleanup costs, the environmental impact and the revenue loss that will be incurred by the operators. There is therefore a need for early detection of pipeline leaks as this will reduce the total volume of oil spilled into the environment and all the associated costs.

1.1 Pipeline Leak Detection Strategies

There are several strategies employed by operators for detecting pipeline leaks and all these approaches have their areas of strength and weaknesses. The most basic approach includes the use of line trackers and monitors to monitor the pipeline Right of way and report any leaks to the operators. While this has a low deployment cost, it does not provide real-time coverage of the line and some sections of the lines are in inaccessible locations thus the leaks can only be discovered by the trackers when it gets to their villages or farmlands. Another approach utilized by operators includes the use of sensors installed along the line to detect the presence of pipeline leaks and report the same to the control room. Table 1.1 presents a summary of key leak detection strategies and their operational characteristics.

Table 1. 1 Characteristics of leak detection strategies **(Fiedler 2016)**

Method	**Typical Min Detectable Leak Rate**	**Time to Detect leak (liquid)**	**Time to Detect leak (gases)**	**Detectable types of leaks**	**False Alarm**
Pressure Point Analysis	> 5%	Short	Long	Spontaneous leaks	High
Mass balance Method	>1%	Long	Very long	Spontaneous and Creeping leaks	High
Statistical Methods	>0.5%	Long	Very Long	Spontaneous and Creeping leaks	Slight
RTTM	>1%	Short	Short	Spontaneous and Creeping leaks	Average
E-RTTM	>0.5%	Very Short	Short	Spontaneous and Creeping leaks	Slight

1.2 Research Gaps

The most significant research gap in literature is the lack of access to reliable pipeline leak data. Operators are not ready to release pipeline leak datasets for research due to the proprietary nature of the datasets and the regulatory implications of these datasets. (Adegboye et al. 2019).

The second research gap is the delays and sometimes inability of the existing leak detection systems to detect low pressure/low volume leaks located far away from the inlet of the pipeline. In this case, the pressure variation generated by the leak dissipates before it gets to the inlet pressure sensors and the flow volume difference is masked by the threshold values and fluctuation associated with the pipeline fluid. This results in the delayed detection of the pipeline leaks as shown in (Baumgarten 2022).

The high cost associated with the development, deployment and management of leak detection systems is another challenge associated with the delays with deployment of the appropriate pipeline leak detection systems.

1.3 Research Questions

The following Research questions have been identified from a review of the literature and the leak detection practice in the industry.

1. How can pipeline leak and no-leak datasets for specific pipeline networks be generated using simulation software.
2. How can leak detection systems be designed to increase their sensitivity to low-pressure leaks that occur far away from the inlet sensors, and which cannot be detected by these inlet sensors.

1.4 Research Aim and Objectives

The aim of this research is the development of a low-cost leak detection system to detect low-pressure/low volume pipeline leaks located far away from the inlet sensors.

The objectives are listed below.

1. Review existing data mining-based pipeline detection systems.
2. Develop a horizontal pipeline model and generate the pipeline pressure profile
3. Generate time series pipeline datasets using PIPESIM and Python for the leak/ No Leak Scenarios.
4. Develop a leak detection architecture for the detection of low pressure pipeline leaks

5. Develop the leak detection algorithm using the developed leak detection architecture.

1.5 Methodology

The following approaches will be used in the execution of the research.

1. Review existing Anomaly Detection methods for Time Series Data.
2. Develop a pipeline model and simulate leak scenarios using PIPESIM and suitable simulation software to acquire pipeline leak dataset.
3. Develop a pipeline dataset-generating model using PIPESIM and Python programming package.
4. Develop the leak detection topology for detecting low volume leaks.

1.6 Significance

1. The PIPESIM simulation software and Python programming will be used to simulate the pipeline network and generate pipeline pressure datasets.
2. The use of this simulation software will enable the generation of pipeline leak datasets which is one of the most significant challenges facing pipeline leak detection research.
3. Deep learning models will be used to develop the leak detection and localization algorithm and the performance of these models will be evaluated using the simulated data.

1.7. Expected Contribution

The expected contributions include

1. A framework for generating both leak and no-leak pipeline datasets for operator pipeline network using PIPESIM and Python software

2. A low cost leak detection system for detecting low volume leaks which would ordinarily be missed by the existing pipeline leak detection systems

1.8 Thesis Structure

This thesis consists of six chapters.

Chapter 1 is an introduction to the research. A brief overview of pipeline leaks is given. The objectives, methodology and significance of this project are also presented.

Chapter 2 includes a literature review of the studies related to pipeline leak detection and time series anomaly detection. The review covers time series data analytics, anomaly detection models, and a brief overview of the simulation software used in this study, PIPESIM. The features, elements and capabilities of the software will form the content of this Chapter.

In Chapter 3, the materials and methods is presented. The experimental processes deployed to answer the research questions which includes simulations utilizing PIPESIM is presented in the chapter.

Chapter 4 is a presentation of the results obtained from the numerical simulations performed in this study.

In Chapter 5, a discussions of the results obtained in the research is presented.

Chapter 6 presents the research contributions made by this work and also provides recommendations for further research.

1.9 Summary

In this Chapter, we presented a brief introduction to pipeline leak detection. The objectives, significance, and methodology related to this project, as well as the structure of this thesis were also presented.

In the next Chapter, a review of the literature will be presented to give a background to the challenge of leak detection and the various time series data analytics and anomaly detection methods. This will include a review of field studies, analytical solutions, experimental methods, and numerical simulation.

Chapter 2

Literature Review

In this Chapter, we present a review of the literature on pipeline leak detection systems and algorithms. The chapter is divided into sections covering pipeline leak detection using Machine learning algorithms, Time series anomaly detection, and the one-class classification algorithms for univariate datasets.

2.0 Low Pressure Pipeline Leak Challenge

The severity of pipeline spills is determined by the volume of oil spilled into the environment. A review of pipeline spills shows that the early response to the spills would have reduced the environmental impact and the financial losses incurred by the companies. The key reasons for the delayed response to the oil spills include the following.

1. The installed leak detection systems are not sensitive to the low-volume leaks located in remote locations far from the inlet and outlet pressure sensors. This challenge is validated by the oil spill which occurred in Alaska’s North Slope. This spill deposited 267,000 gallons of thick crude oil over two acres in the Prudhoe Bay production facilities. The spill went undetected for as long as five days before an oilfield worker detected the acrid scent of hydrocarbons while

driving through the area. The investigation revealed that the pressure of the leaking oil gradually expanded the hole to a quarter- or half-inch wide. Most of the oil seeped beneath the snow without attracting the attention of workers monitoring alarm systems”. Among the reasons for the delayed detection was the fact that the leak was "smaller than our system would detect” and that the normal fluctuations of oil flow in this particular pipe could have masked warning signals (Baringer 2006).

2. The Operator’s personnel are not trained to understand the leak signals and alarms generated by the system. The Alberta Energy Regulator (AER) reports concluded “that company personnel responsible for leak detection were not sufficiently trained or simply failed to recognize that a leak was occurring until several days after the leak had started" (Bickis 2017).

3. The minimum leak detectable by the installed LDS is a percentage of the liquid flow in the pipeline such that for high capacity pipelines, the minimum leaks that can be detected will already be too large for the environment to handle (Baringer 2006).

4. The unwillingness of the operators to shut down the lines even when they have received the notification of pipeline line leaks on their lines. This was the case with Husky Energy which leaked 225,000 barrels. The company waited for 10 hours before shutting the line even after receiving the leak alerts. In 2010, Enbridge pipeline operators took 17 hours to shut down a pipeline after receiving the oil spill alerts. This delay resulted in the spill of about 3.3 million

liters of crude which flowed into Michigan's Kalamazoo River (Fitzpatrick et al. 2015).

5. The lack of infrastructure for the regulators to monitor the leak detection systems of the operators. This forces the regulators to rely on the estimates provided by the operators which in most cases are contested. A typical case is the AER which does not track how leaks are discovered along the pipelines it oversees in the province. It relies on companies to submit that information in spill reports. A review by the Alberta regulator observed that improper leak detection contributed significantly to 8 of the 23 pipeline spills it investigated in 2013. Several other regulators in different countries rely on third parties for reporting cases of pipeline spills (Bickis 2017). The provincial government of Saskatchewan only requests a spill report but does not require the details of the leak detection. In Nigeria, the National Oil Spill Detection and Response Agency (NOSDRA 2022) rely on third-party information and provides estimates of the spills. It has a citizen reporting system where citizens can report identified leaks to the organization which then undertakes steps to estimate the volume of oil spilled from the lines. NOSDRA.

With over 1600 individual leaks in the US alone and a total volume of over 11 Million gallons spilled since 2010, pipeline spills continue to be a very critical area of research for the Industry, the environment, and the government. Over 700 leaks were reported in North Dakota between 2010 and 2016 releasing an average of 5000 gallons per incident. The largest of these spills (September 2013) which was also one of the largest onshore oil spills in the U.S was not detected by the Leak Detection Systems of the operator. Over 865,000 gallons of crude oil spilled into a wheat field contaminating about 13 acres and all the remote monitors missed it. It took the vigilance of the owner of the field to detect the spill. Pipelines have been known to leak over 1 million gallons in one single leak with these leaks going undetected in some cases due to either the sensors' insensitivity or the operators' unwillingness to shut the lines. While the operators have leak detection systems installed, some of these LDS are not able to detect leaks that fall below their minimum detectable range. This minimum is often a function of the line's total fluid-carrying capacity, which can run to several hundred thousand barrels per day. Some of these leaks which occur far away from the inlet stations go undetected by the operators for several days or weeks and are detected by the locals in the vicinity of the lines. Unfortunately, at the time of their detection, the spills would have done major damage to the environment around which they occurred. The Operation of old and aging pipeline infrastructure is another justification for the installation of a robust leak detection system. The companies are emboldened to continue production even when their alarms systems indicate the presence of leaks because the regulators do not have access to the leak monitoring platforms of the operators. The use of satellites for image capture and processing is a new approach being deployed for detecting these low-pressure leaks however, in cases like North Dakota with a lot of snow and ice cover and for underground pipelines, this application of satellite for leak detection will not be as effective as is desired. (Satelytics 2016). Table 2.1 shows some pipeline spills and the associated delays with their detection.

Table 2. 1 Shows some pipeline oil spills and the delays associated with their detection

Leak	State	Duration to Detection	Findings	Literature Source
Pipeline Rupture Discovered After 143 Days and Discharge of 29 Million Gallons	North Dakota	143 days (More than 4 Months)	The company gave misleading and incomplete statements to the government about the duration and size of the spill and also did not install reliable LDS and even when they got the indications of a leak, the continued production.	(Department of Justice, 2021).
Pipeline carrying oil ruptured in a farmer's field, spilling an estimated 20,600 barrels over a seven acre area.	North Dakota	Leak went undetected by the company until a farmer drove his combine into a sodden field of oil.	Leak was detected by a farmer.	(Coleman 2014)
Husky spilled 225,000-litre in Saskatchewan	Saskatchewan.	10 hours after receipt of alerts.	The company waited for 10 hours before shutting the line even after receiving the leak alerts	(DeBofsky et al. 2020)
Enbridge pipeline spilled about 3.3 million litres of crude which flowed into Michigan's Kalamazoo River.	Michigan	pipeline operators took 17 hours to shut down a pipeline after receiving the oil spill alerts.	pipeline operators took 17 hours to shut down a pipeline after receiving the oil spill alerts.	(Fitzpatrick et al. 2015)
The oil spill occurred in Alaska's North Slope. This spill deposited 267,000 gallons of thick crude oil over two acres.	Alaska	5 Days. Was detected by an oil field worker due to the acrid smell from the environment	The leak was "smaller than our system would detect" and the normal fluctuations of oil flow in this particular pipe could have masked warning signals.	(Barringer 2006)

2.1. Pipeline Transport of Oil and Gas Products

Pipelines are the most economical means of transporting huge volumes of petroleum products across large geographical locations, however, these lines come with the possibility of pipeline leaks which are one of the most devastating environmental challenges faced by communities around the pipeline right of way (RoW). Typical impacts faced by these communities include the loss of their livelihoods, especially the fishing and farming communities, the local fauna and flora are also impacted by the leaks. The resulting flames from these pipeline leaks in the event of a fire results in the pollution of the environment resulting in expensive cleanup costs and the associated fines and economic losses to the companies. There are almost 3 million kilometers of pipelines currently in service globally, (Akinsete and Oshingbesan 2019). With over 830,000 km of oil and gas pipelines, North America has the longest oil and gas pipelines. Of these numbers, 154,200.9 km are pipelines that carry crude oil, 103,106.3 km of the pipeline transport petroleum products, 495,555.3 km of the pipelines is responsible for natural gas transport and 81,290.0 km is reserved for NGL. The US accounts for 41% of the global pipeline infrastructure. (Idachaba and Minou, 2021). The integrity of the pipelines is impacted by material defects, corrosion, abrasion, and other third-party intrusion activities. These activities lead to the loss of containment resulting in an oil spill or a pipeline leak. (Wang and Duncan 2014; Bolotina et al. 2018).

Costs associated with pipeline accidents in the US alone are close to $7 Billion in property damages, spilling millions of barrels of oil into the environment. Figure 2.1 shows the oil spill occurrences in the US from 1986 to 2020. (Richard 2013).

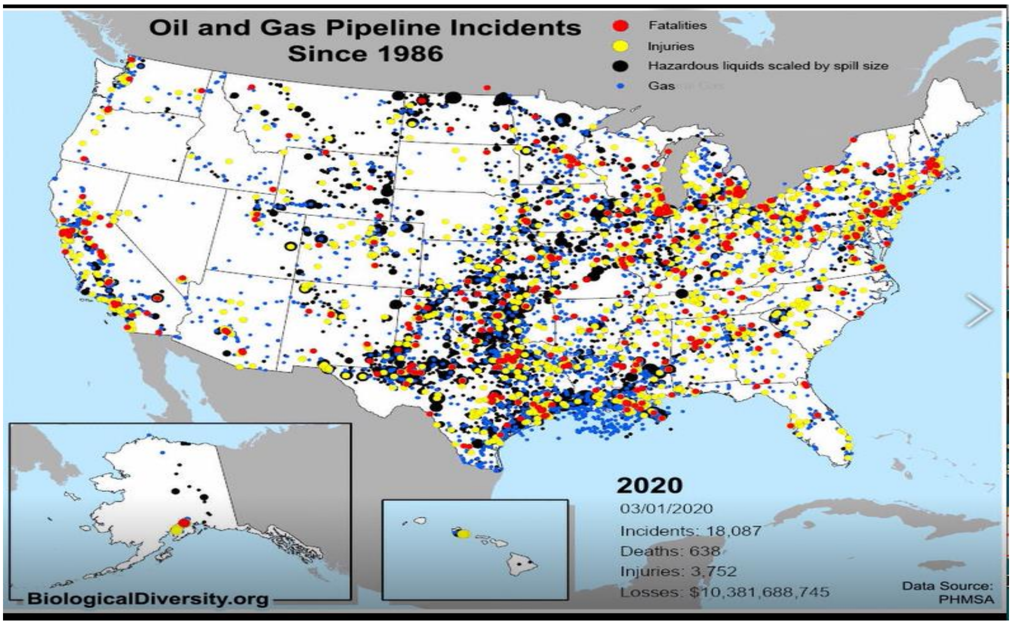


Figure 2.1 USA pipeline leak locations and the impact. (Richard 2013)

There has been over 18000 incidents leading to over 600 deaths 3752 injuries and a loss of over $10 Billion. The Kalamazoo river oil spill resulted in an over $800 Million loss in property damage with more than 840,000 gallons of crude oil spilled into the Kalamazoo River. (Lena 2012).

2.2 Pipeline Leak Lifecycle

The severity of a pipeline leak is determined by the nature of the pipeline failure, the leak size, the leak identification, and eventual localization. The pipeline failure can either be a corrosion failure in which case the failure can be a pin hole size failure, a crack occasioned by third party activities, or it can be pipeline rupture where there is a separation of the pipeline section. The volume of fluid flow to the environment varies with the type of pipeline failure, The leak size refers to the dimension of the leak point and this determines the flowrate from the pipeline. This range from the pin hole size and increases as the dimension of the leak increases. The larger the leak size, the higher the volume of oil spilled into the environment in the event of a pipeline failure. These show that whenever there is a leak, the total spill volume is dependent on the time of the leak detection and the commencement of containment activities. Regardless of the type of pipeline failure, the ability of the leak detection system to both determine the presence of a leak as soon as it starts and determine the location are the two most significant features of any leak detection system. The pipeline leak lifecycle is shown in Figure 1.1. When there is a pipeline failure, the pipe begins to spill its products into the environment until the leak is detected. The detection speed is a critical feature in the leak detection system as the state of the sensors and the size of the leak affect the detection speed. The parameters for detecting the occurrence of a leak include pressure sensors, flowrate sensors, and temperature sensors. For sensors with erroneous readings or desensitized sensors, they may not be able to detect minute changes in the flow parameter in the event of a small leak. This is further reinforced if the leak is in a remote location far away from the sensors. The leak signal may be too small to be noticed by the sensors when it arrives at the sensor location. If the leak point is in a remote location, the operator may not be able to locate the site until the leak spreads to more visible locations. This was the situation with the North Dakota leak reported in Williams County in North Dakota in 2022. The delayed detection of the spill resulted in an actual spill of 1.4 million gallons which is much more than the 8,400-gallon earlier reported by the company. This rise in the actual volume was due to the delays in detecting the spill as it went on for over a month undetected. Reasons for the delayed detection include the fact that the leak point was in a remote location and was detected by one of the local farmers in a small town 90 miles from Minot the nearest big city. The impact of early detection is also validated by the largest oil field spill in North Dakota which went on undetected for five months in 2014 and 2015 and spilled 29 Million gallons of produced water near Williston, N.D. The estimate initial given was about 70,000 barrels over 10 days, but the final investigations showed that the spill had been on for over 5 months resulting in the 29 million gallons spilled (Baumgarten 2022)

The oil spill volume continues to increase until the appropriate oil spill response is deployed to contain the spill. Some of these actions range from shutting down the line to diverting flow to other lines. The rate of discharge is reduced after the remedial action, but the spill continues until the line is completely discharged. There is therefore a need for early detection of very small spills to minimize the volume of oil spilled into the environment.

Pipeline Leak Detection Regulation

Several countries in a bid to better manage the risks associated with pipeline transport of petroleum products put regulations in place to manage pipeline safety and operations. These regulations and their countries of origin include: (Idachaba and Minou 2021)

- The TRFL which is Germany's Technical Rule for Pipelines
- Three API standards for the USA
  - Computational pipeline monitoring for liquids. API 1130
  - Variable uncertainties in pipelines and their effects on leak detection performance. API 1149
  - Performance criteria for leak detection systems API 1155 which has been replaced by API 1130
  - Transport of hazardous liquids via pipeline 49 CFR 195
- Canada - CSA Z662, focuses on oil and gas pipelines

2.3 Pipeline Leak Detection Methods

Pipeline leaks are due to a loss of containment and the release of the pipeline content into the environment. These leaks are associated with a loss of pressure in the pipeline and the volume of the oil leaked out of the pipeline depends among other on the size of the leak orifice and the pressure in the pipeline. The flowrate of the leak is impacted by the pressure in the line This flowrate increases when pressure is high and decreases when pressure is low. This is why the U.S. Department of Transportation Standards for emergency response for pipeline facilities, recommends the isolation of the leak point by shutting both the upstream and downstream valves to reduce the pressure in the line and as this will reduce the leak flowrate. (U.S. Department of Transportation Standards 2010).

There are two broad classifications of leak detection systems, these include continuous and noncontinuous systems. Often, the continuous and non-continuous systems are used together (Romero-Tapia et al. 2018). The non-continuous systems comprise the following activities (Baroudi et al. 2017)

1. Inspection by helicopter,
2. smart pigging,
3. even tracking dogs, and
4. RoW monitoring and patrol

These approaches are often reactive as they are usually triggered by a drop in pressure indicative of a leak event or scheduled routine surveillance.

The continuous method can be further classified into internal and external based systems depending on the location of the sensors. (Romero-Tapia et al. 2018; Baroudi et al. 2017).

The external systems include the use of

1. fibre optic cables,
2. acoustic systems,
3. semi-permeable sensor hoses, and

4. video monitoring.

Internal systems include:

1. pressure point analysis,
2. Mass balance method,
3. Statistical systems,
4. RTTM-based systems, and
5. Extended RTTM.

Pipeline Leak detection systems can also be broadly classified as being either hardware-based or Software based leak detection systems. (Murvay and Silea 2012; Vrålstad et al. 2013; Scott and Barrufet 2003). For hardware-based systems, sensors are installed on the lines for the purpose of data acquisition. These methods which are also referred to as the exterior methods are costly to deploy and maintain and are subject to vandalization and hardware failure. The systems also require data transmission capacities which in some cases would require the use of licensed frequency bands for long-range transmission. (Idachaba et al. 2014).

The software-based systems rely on pipeline models and measurements such as (pressure, temperature, flowrate, and differential pressure) which are fed into specially developed leak detection and localization algorithms for the detection and localization of pipeline leaks. These methods are cheaper to deploy as they rely on pipeline production data. Inspite of the low cost associated with the software-based methods, it is faced with the challenge of a lack of pipeline leak datasets (operators are not ready to release these datasets) which makes it difficult for researchers to develop leak detection algorithms using field datasets. A summary of the different pipeline leak detection methods are shown in Figure 2.2.

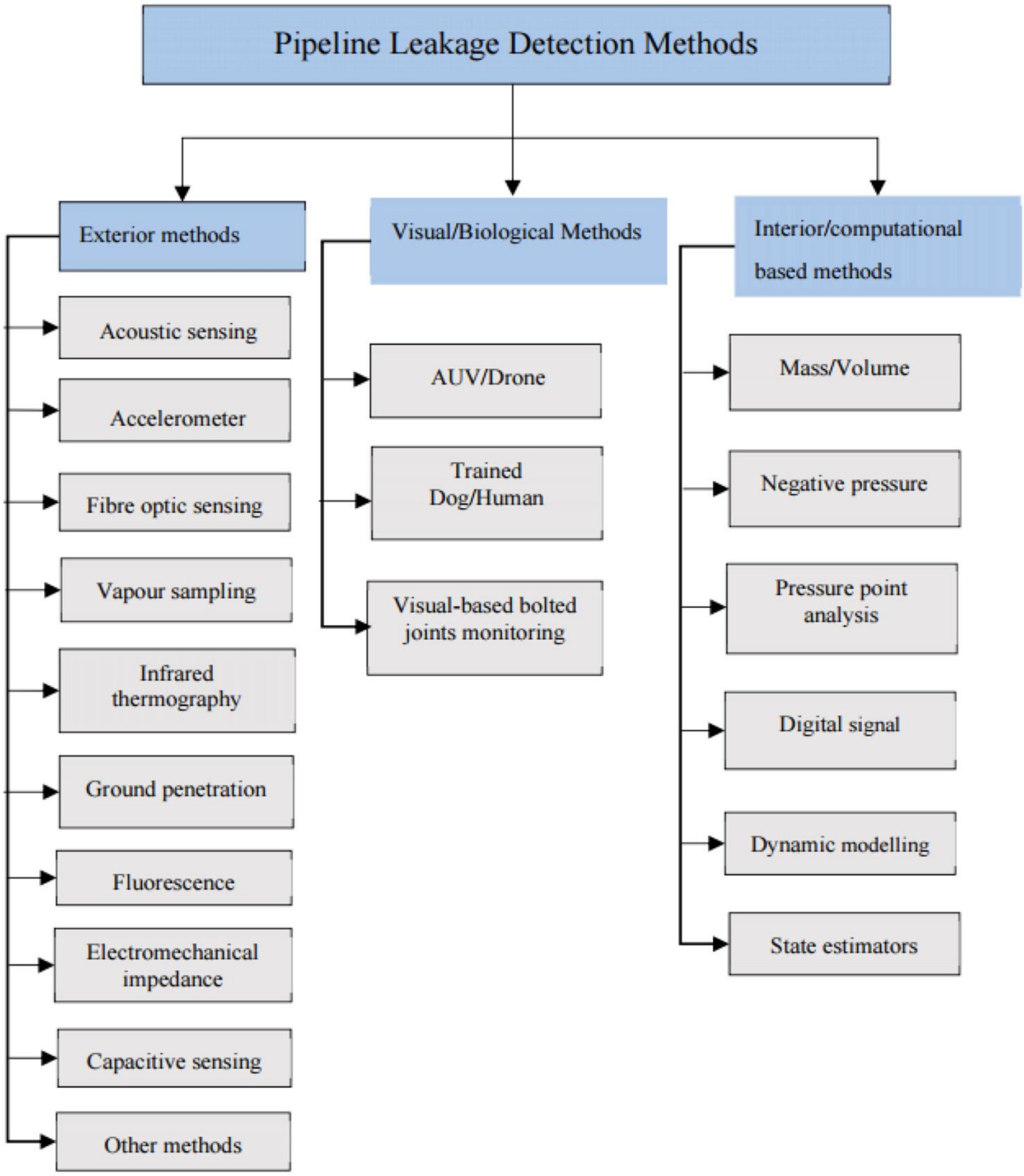


Figure 2.2 Leak Detection Methods (Adegboye et al. 2019).

As shown in Figure 2.2, Pipeline leak detection methods can also be broadly classified into three groups which are the exterior, visual/biological and the interior/ computational based methods. The exterior methods are largely hardware centric and the require the installation of sensors on the pipeline for data acquisition. The visual inspection method is one of the most popular methods for leak detection.

Other exterior methods comprise the use of acoustic leak detection systems, infrared thermography, ultrasonic methods, or electromagnetic techniques. (Scott and Barrufet 2003; Adegboye et al. 2019). These systems enable the precise location of the leaks but have the disadvantage of being very expensive and taking very long durations before the leak can be located. The accuracy of detection also for these systems rely on the reliability of the sensors because when the sensor sensitivity begins to drift, the detection accuracy will be affected. The analysis of transient pressure waves is another method for the detection and localization of pipeline leaks. The occurrence of a leak results in the generation of pressure wave which travels throughout the line and are detected by the pressure sensors at both ends of the line. The data received by both sensors are analyzed to predict the location of the identified leaks. (Misiunas et al. 2005; Covas et al. 2005).

The interior or computational based technique which is also software driven utilizes pipeline pressure and flowrate data and a model of the pipeline to detect any variation between the predicted and the measured values. The system relies on algorithms that generate alarms when these difference in values exceed a given threshold and calculates the leak location and magnitude.(Begovich et al. 2012). This system is also referred to as the internal method due to the reliance on pipeline data for the leak detection analysis

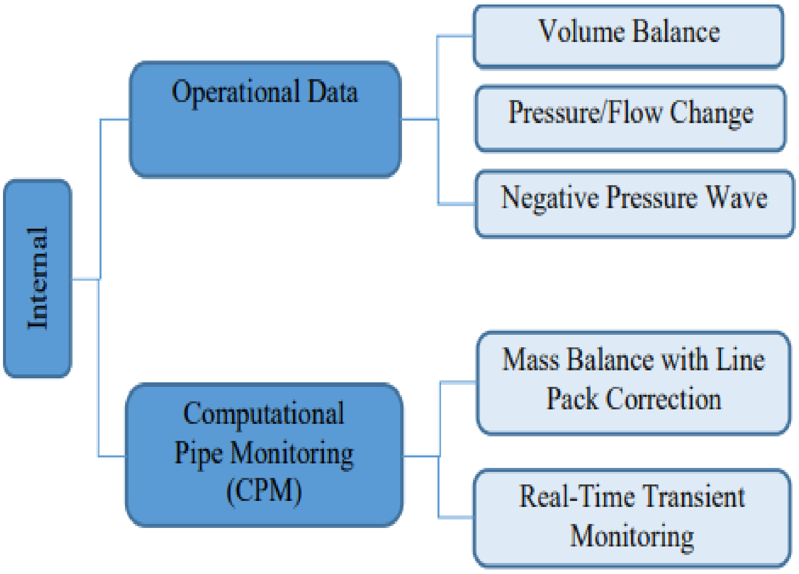


Figure 2.3 Internal Leak detection method. (Baroudi et al. 2019)

From Figure 2.3, the Internal systems utilize either the operational data or the Computational Pipeline Monitoring (CPM) approach. Techniques which utilize the operational data include the volume balance technique, the pressure /flow change technique or the negative pressure wave technique. The Computational Pipeline Monitoring on the other hand utilizes either the mass balance with line pack correction or the Real Time Transient Monitoring approach. The internal method relies on field sensors to monitor the operational and hydraulic conditions of the pipeline, e.g., measurements of the flow, pressure, and temperature. These real-time parameters are compared with the normal working parameters of the pipeline which are determined either manually by pipeline controllers or based on sophisticated algorithms and hydraulic models. (Baroudi et al. 2019)

A difference between the measured and predicted operational parameters indicates a leak. The remote field sensors installed on the pipeline monitor the lines continuously and send the data to a centralized monitoring station, where the data undergoes filtering, signal processing and is passed on to the leak detection and localization algorithms to both detect the presence of the leaks and also identify its location.

The data acquisition methods from the sensors include the following.

1. **Negative Pressure Wave (NPW):** Negative pressure waves are created by sudden drops in pressure caused by leaks, and they propagate in both directions from the leak point. A critical challenge of this system is that it cannot differentiate between leaks and normal operations, and this results in false alarms.
2. **Volume Balance:** This is the volume differential between the incoming and outgoing volumes. Volume balance can detect catastrophic failures; however, its usage is rare due to its limited performance.
3. **Rate of pressure/flow change**: Leaks are associated with sudden changes in pressure. However, sudden pressure variations can also be due to transient conditions of the pipeline. Filtering techniques and suitable algorithms are used to differentiate between leaks and operations-induced pressure changes. Pressure waves also dampen out as they traverse a longer length and thus additional pressure sensors need to be installed along the pipelines.
4. **Computational Pipeline Monitoring (CPM)** This method detects anomalies in pipeline operating parameters, and this is accomplished using the Mass Balance with line pack approach and the Real Time Transient Modelling method. (Baroudi et al. 2019)

Current trends in pipeline leak detection and localization methods further identify two main categories for pipeline leak detection under the internal method using real-time transient modelling and pipeline data. These are the signal-based methods and the model-based methods. These methods are based on the steady-state models of the pipelines, they can predict unknown parameters and determine the occurrence and localization of leaks more accurately.

The pipeline-model-based method predicts pressure distribution along the pipeline and locates the leak through the pressure and flowrate signal on both ends of the pipeline. The most popular and widely used pipeline model-based method includes the pressure gradient (PG) method and the average friction coefficient (AFC) method. (Chuanbo et al. 2018)

The pressure gradient PG method ignores the influences of friction coefficient, temperature, pipe diameter and other factors on the pressure distribution along the pipeline and considers the pressure to be linear. The average friction coefficient (AFC) method on the other hand assumes the friction coefficient and pipe diameter to be constant. (Chuanbo et al. 2018)

A summary of the classification of pipeline leak detection methods is shown in figure 2.4.

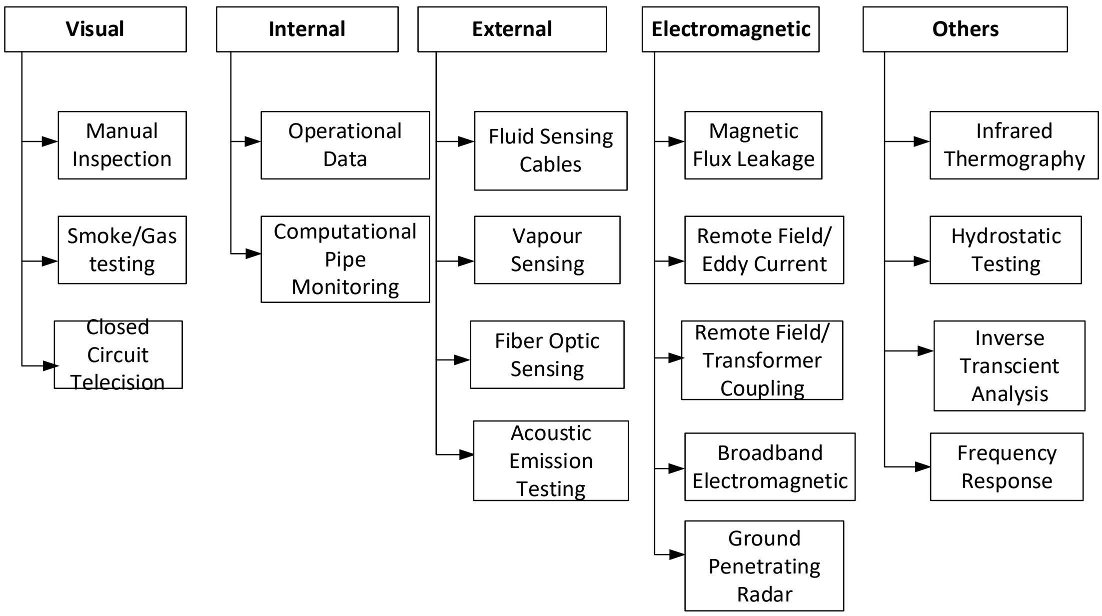


Figure 2.4 Classification of pipeline leak detection systems (Baroudi et al. 2019)

Literature has shown the availability of multiple leak detection strategies; however, the evaluation of these different Leak Detection Systems (API 1130) relies on four key parameters (Taylor and Hamidreza 2019).

1. Accuracy: the LDS should be able to calculate leak size and leak location accurately. This is quantified as the maximum distance allowed between the estimated location and the actual location, as well as the maximum variation allowed between the estimated and actual leak size.
2. Reliability: the LDS should correctly display any real alarms and not report any false alarms. This should be quantified in the number of false alarms acceptable.
3. Robustness: the LDS should be able to operate in non-ideal environments, such as when sensor input equipment fails to provide data. This should be quantified in % availability.
4. Sensitivity: the LDS should be able to detect small leaks and detect them quickly. This is quantified in absolute flow rate terms because a relative change can be misleading.

The performance criteria of the API 1155 which was replaced by the API 1130 is shown in Table 2.3

2.4 Pipeline Modelling

The scarcity of pipeline datasets is one key challenge facing research in pipeline leak detection research. Operators are unwilling to share the datasets and researchers cannot fund data acquisition from actual pipelines. This justifies the use of hypothetical data generated by the transmission pipeline model (Sukarno et al. 2007).

The transient pipeline flow model provides the foundation for pipeline simulation and modelling. The basic equations governing this model include the continuity, the momentum, the energy equations and the equation of state. (Chaczykowski 2010).The continuity equation expressed in the equation 2.1 focuses on the conservation of mass principle. It requires that the difference in mass flow into and out of any section of the pipeline is equal to the rate of change of mass within the section. This is shown in equation 2.1

$$\frac{d(\rho)}{dt} + \rho \frac{\delta(V)}{\delta s} = 0 \qquad 2.1$$

In this equation, ρ = density, t = time, V = flow velocity and s = pipeline location coordinates

The conservation of momentum equation is represented in equation 2.2 :

$$\frac{d(V)}{dt} + \frac{1}{\rho}\left(\frac{\delta P}{\delta s}\right) + fs = 0 \qquad 2.2$$

In this equation, V = the flow velocity v, t = time, P = pressure , s = pipeline location coordinates and fs = pipeline friction.

The conservation of energy principle is represented in equation 2.3.

$$\frac{dh}{dt} - \frac{1}{\rho}\left(\frac{dp}{dt}\right) - I_L = 0 \qquad 2.3$$

In this equation, h = enthalpy , t= time, ρ = density, P = pressure , I_L= specific loss performance L

The simulation of a leak is accomplished by introducing a branch pipe of a given diameter on the main pipeline. This branch pipe can be located at any point on the main line with the leakage rate made variable. The variable leakage rate enables the study of different leak types on the main pipeline. This pipe model is represented schematically in Figure 2.5. In this Figure, *D*1 is the distance between the leak point and upstream pressure sensor and *D*2 is the distance between the leak point the and downstream pressure Sensor while S1 and S2 are the inlet and outlet sensors respectively. The fluid flow is represented by the direction of the arrows which are from the Pin to the Pout.

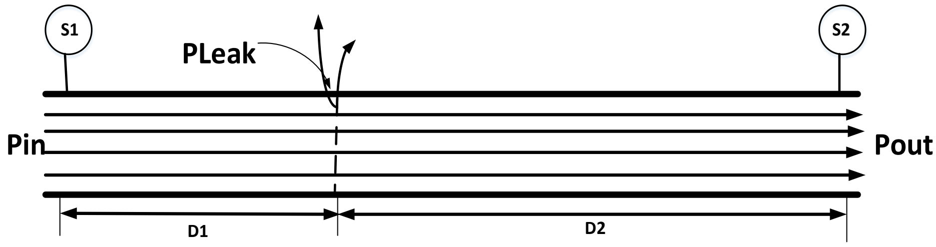


Figure 2.5 Pipeline with the leak at distance D1 from the inflow section

Pipeline leaks can be characterized by a sudden drop in the pressure at the leak point after which the pressure rises to a value below the initial pre-leak value. The leak point is bounded by the upstream point which is the pipeline portion between the inlet and the leak point while the downstream portion is the pipeline section between the leak point and the outlet of the pipeline. This pipeline model can produce sufficient data for different leak conditions and the data will be used for training the leak detection model.

The mass flow rate into the system = Mass flow rate out of the system. For compressible fluid applying conservation of mass,

$$\rho_{in}A_{in}V_{in} = \rho_{out}A_{out}V_{out} \tag{2.4}$$

A = Cross sectional area of the pipe

V = Velocity of fluid

ρ = Fluid density

The pressure at the inlet and outlet of the pipeline is represented by equations 2.5 and 2.6

$$Pin = Pout + Pleak + Ploss \tag{2.5}$$

$$Pout = \beta Pin - PLeak - PLoss \tag{2.6}$$

Where β=Pressure Loss Factor due to the length of the pipeline

$PLoss$ = Pressure loss due to wax buildup in the pipeline (This is 0 for new pipelines)

$PLeak$ = Pressure loss due to leak.

From the diagram in figure 2.5,

$D1$ = Distance between Leak point and Upstream Pressure Sensor

$D2$ = Distance between Leak point and Downstream Pressure Sensor

The pressure at the upstream end is p, and at the downstream end the pressure has fallen by Δp . This value becomes = $p - \Delta p$.

The driving force due to pressure (F = Pressure x Area) can then be written as:

driving force=pressure force at input-pressure force at the output as shown in equation 2.7

$$pA - (p - \Delta p)A = \Delta pA = \Delta p \frac{\pi d^2}{4} \qquad 2.7$$

The retarding force is the force due to the shear stress by the walls = sheer stress x area of pipe wall. The nature of the fluid flow in the pipeline is represented by the Reynolds number.

The Reynolds number formula is expressed by equation 2.8. **(**Smits and Hultmark 2014)

$$R_e = \frac{\rho V D}{\mu} \qquad 2.8$$

Where, μ = Fluid viscosity and

Reynolds number formula is used to determine the velocity, diameter, and viscosity of the fluid.

The kind of flow is based on the value of Re

1. If Re < 2000, the flow is called Laminar
2. If Re > 4000, the flow is called turbulent
3. If 2000 < Re < 4000, the flow is called transition. This parameter will be utilized in the modelling of the petroleum pipeline.

The model shown in figure 2.6 for a pipe section can be extended to cover the entire pipeline network and also extended to enable the detection of multiple pipeline leaks from pipelines. The installation for the entire pipeline network is shown in figure 2.6.

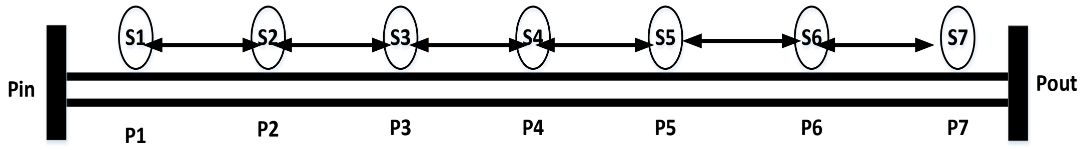


Figure 2.6 Pipeline network modelling for leak detection

The pipeline section in figure 2.6 comprises of sensors installed at specific points. The sensors are represented by S1 to S7 while the pressure values at the different sensors are represented by P1 to P7. The goal of the simulation of this pipeline network is to determine the minimum distance for the installation of sensors which can provide end-to-end leak detection from multiple leak sources.

Once the data has been generated from the pipeline transmission model, the machine learning pipeline shown in figure 2.7 will be utilized to develop the leak detection model which will be deployed for the detection of leaks from the pipeline. The data will also be used in developing the leak localization module.

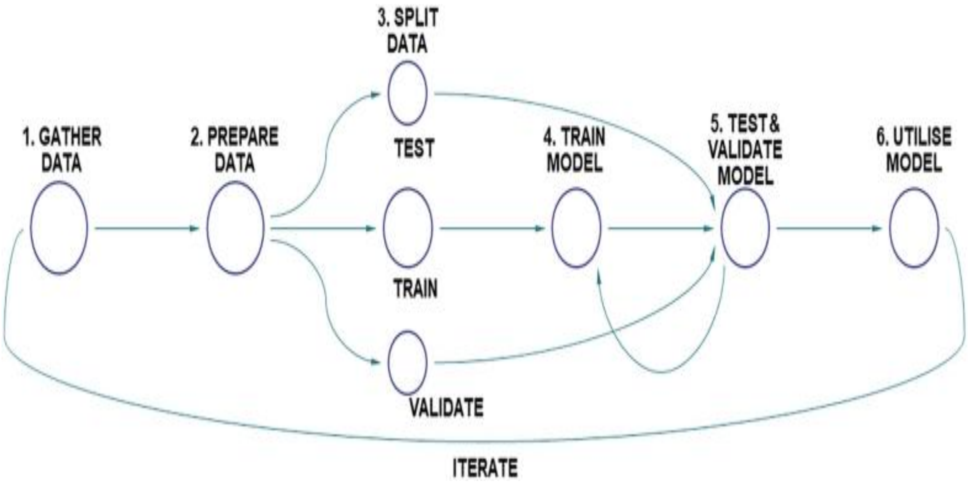


Figure 2.7 Machine Learning pipeline. (Akerkar 2019)

The machine learning pipeline is the processes utilized in developing machine learning models. The data is acquired and cleaned to remove erroneous entries or missing values.

After the data preparation, it is split into training and testing data. Some applications classify the data into training, testing and validation. The model is exposed to only the training data during the training phase. After the training, the model is exposed to the testing data to test and evaluate performance of the data. The model can also be subjected to the validation data after which it deployed if it meets the minimum benchmark criteria. Once the model has been finalized, tested, and validated, it is deployed to monitor the pipeline and develop a database of the pipeline sensor readings. The model will be able to able to learn from the data and optimize itself thereby increasing both its accuracy and speed of detection. The continuous data acquisition leads to the development of the Real Time Transient Model approach for leak detection.

2.4.1 Governing Equations

1. Darcy Weisbach Equation (Renata et al. 2020)

For determining the head loss along the pipeline

$$h_L = f\,\frac{LV^2}{D\,2g} \tag{2.9}$$

Where h_L: load loss (m); f: friction coefficient; L: pipe length (m); D: pipe diameter (m); V: flow velocity (m/s); g gravity acceleration (m/s^2).

2. Reynolds number (Smits and Hultmark 2014)

For determining the flow regime shown previously in equation 2.8

3. Colebrook equation (Colebrook 1939)

For determining the friction factor of the pipeline under turbulent flow conditions equation 2.10 is applied.

$$\frac{1}{\sqrt{f}} = -2log10\,\left(\frac{\varepsilon/D}{3.7} + \frac{2.51}{Re\sqrt{f}}\right) \tag{2.10}$$

4. Moody Charts (Moody and Princeton 1944).

Shown in figure 2.8, the Moody charts are for determining the friction factor of pipelines. The x axis is the Reynolds number while the y axis is the friction factor

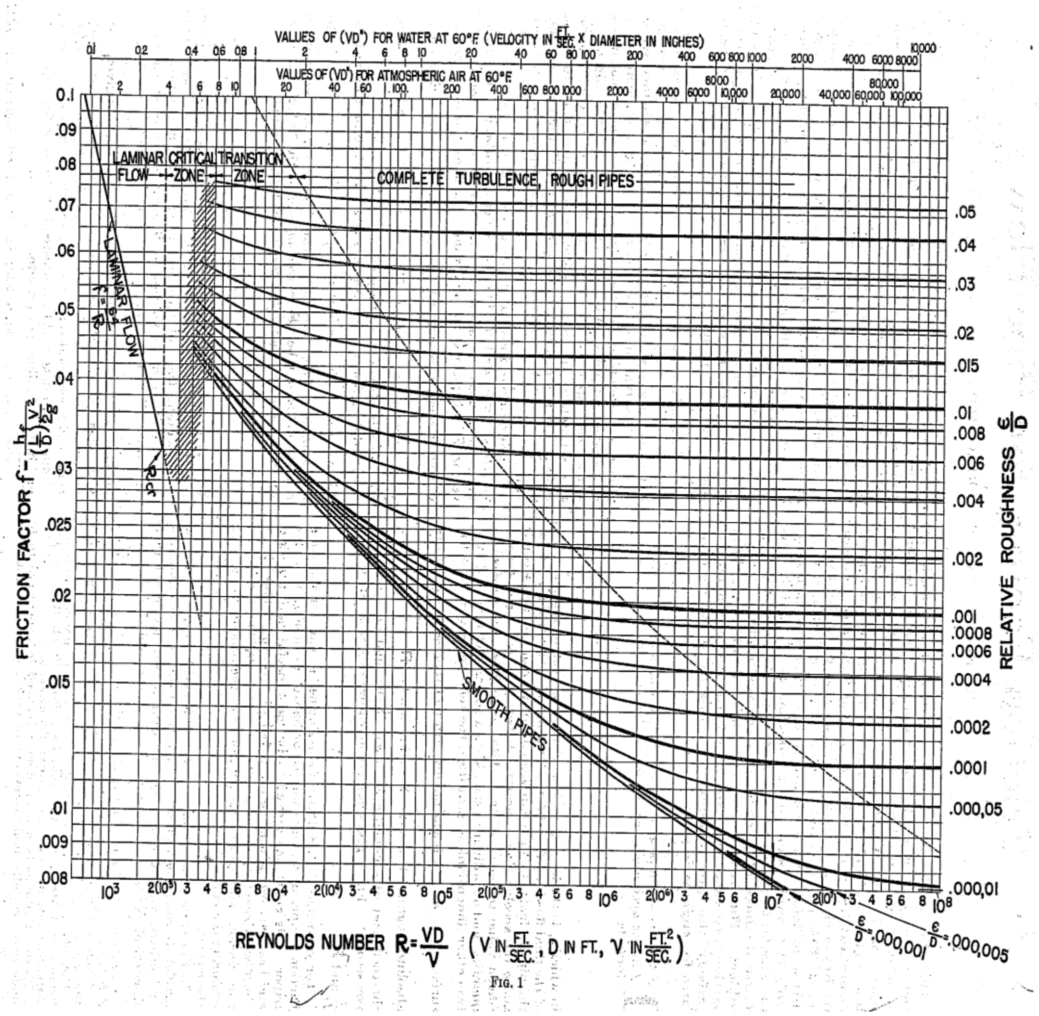


Figure 2.8 Moody Diagram (Moody and Princeton 1944).

2.4.2 Pipeline pressure profiles

The pressure profile shows the pressure values along the length of the pipe. These profiles are discussed under the single and multiple leak conditions.

(a) Single leak conditions

Under single leak conditions, the pipe experiences just one leak . The pressure drop due to the leak results in the generation of negative pressure waves which are transmitted to the sensors located at both the inlet and the outlet of the pipelines. These variations must be outside the pipeline pressure envelope for them to be classified as a leak. Other pipeline activities that can generate these types of pressure variations include a sudden shutting down of the pipeline or an increase/decrease in the inlet pressure of the pipeline. In these cases, pressure fluctuation is expected and the sensor reading can be accurately classified as being due to the known pipeline inlet pressure variations. For cases where these variations exceed the operating envelop while the inlet pressure is retained at the normal operating values, the pressure variations can be classified as a leak occurrence and the required algorithms can be used to locate the leak point

The pipeline pressure profile is shown in figure 2.9. The colors shown in the figure indicate the different pressure profiles present in the pipe. Table 2.2 presents the description of the pressure profiles.

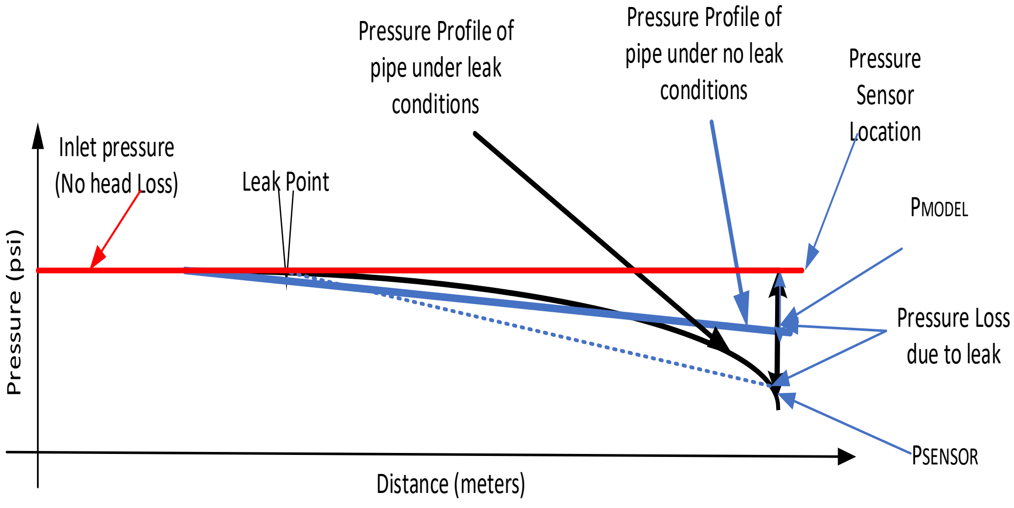


Figure 2.9 Pressure profile for single leak condition

The leak distance from the sensor to the leak point is shown in figure 2.10. Since the sensor position is known, the flowrate is used to determine the distance the negative pressure would have travelled before getting to the sensor.

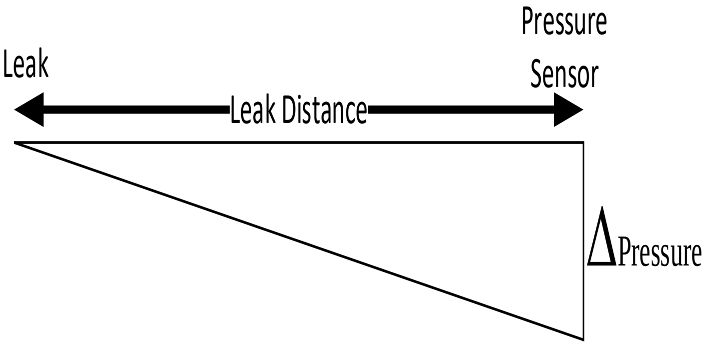


Figure 2.10 Leak Localization

Table 2.2 Pressure profile descriptions

S/N	Line Colour	Pressure Profile
1	Red	This line represents the pressure profile for the pipeline without any head loss. This is an ideal condition as there will always be head loss due to the friction of the pipe walls.
2	Blue	This line represents the pressure profile for the pipeline without any leaks. This is the loss that occurs due to the normal flow of the pipeline. The rate of decay is a function of the inlet pipe and the state of the pipeline's internal walls.
3	Black	This line represents the pressure profile for the pipeline under a leak condition. The leak occurs before the sensor and the negative pressure is transmitted to the sensor and read by the sensor as P_{SENSOR}
4	Dotted Line	This line represents a straight line from the leak point to intersect the leak profile line at the sensor point. This line is used to help determine the leak location

(b) Multiple Leak conditions

For multiple leaks on a pipeline, the analysis done for the single leak is cascaded together for the multiple leak condition. In this case, the pipeline is broken down into segments. The segments are determined by the leak point and the downstream pressure sensor. The inlet pressure changes whenever a leak occurs upstream and this is factored into the equation used for the second segment. This is shown in figure 2.11.

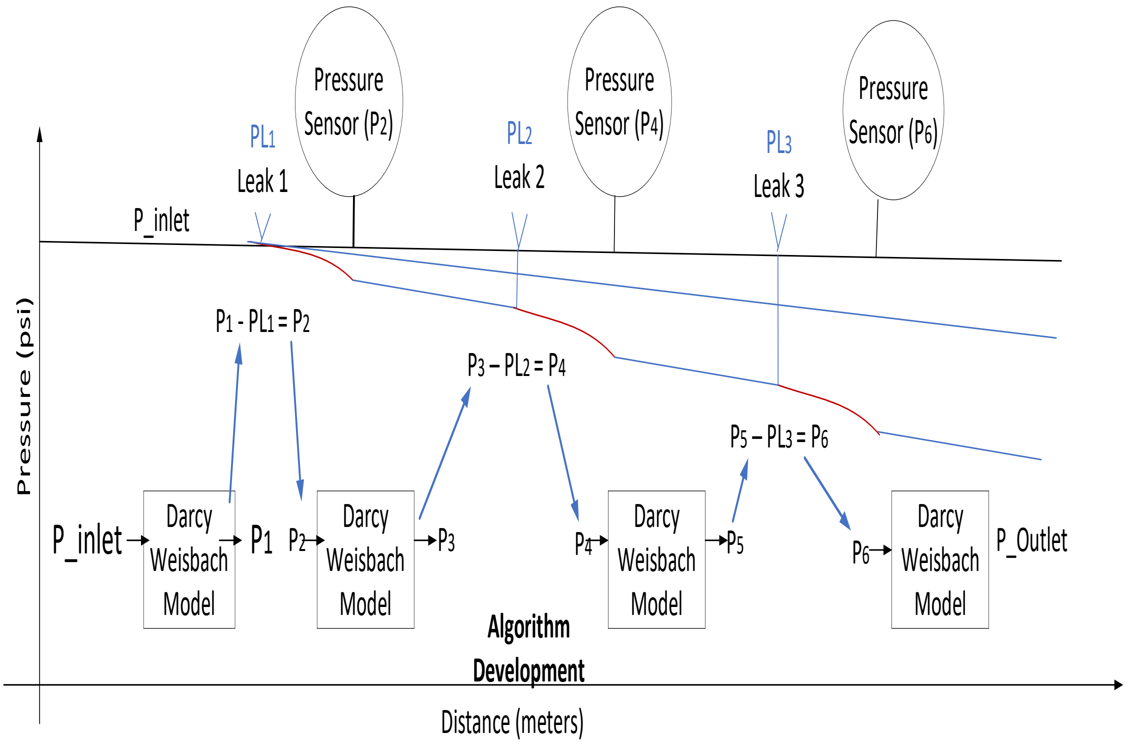


Figure 2.11 Multiple pipeline leak pressure profiles

2.4.3 Correlation of pipeline pressure to time series data

A time series is defined as a sequence of pairs shown in equation 2.11

$$T = \big((p_1, t_1) \dots \dots \dots .. (p_n, t_n)\big)(t_1 < t_2 \dots \dots . < t_n) \qquad 2.11$$

where each pi is a data point in a d-dimensional space and each ti represents the time stamp at which the corresponding pi occurs. (Wang et al. 2013)

The pipeline pressure profile is measured as the pressure along the pipe from the inlet to the outlet. This profile is represented in a graph of pressure vs pipeline length. In converting this to time series, the pressure at each of the sensors is monitored and plotted against time as shown in figure 2.12.

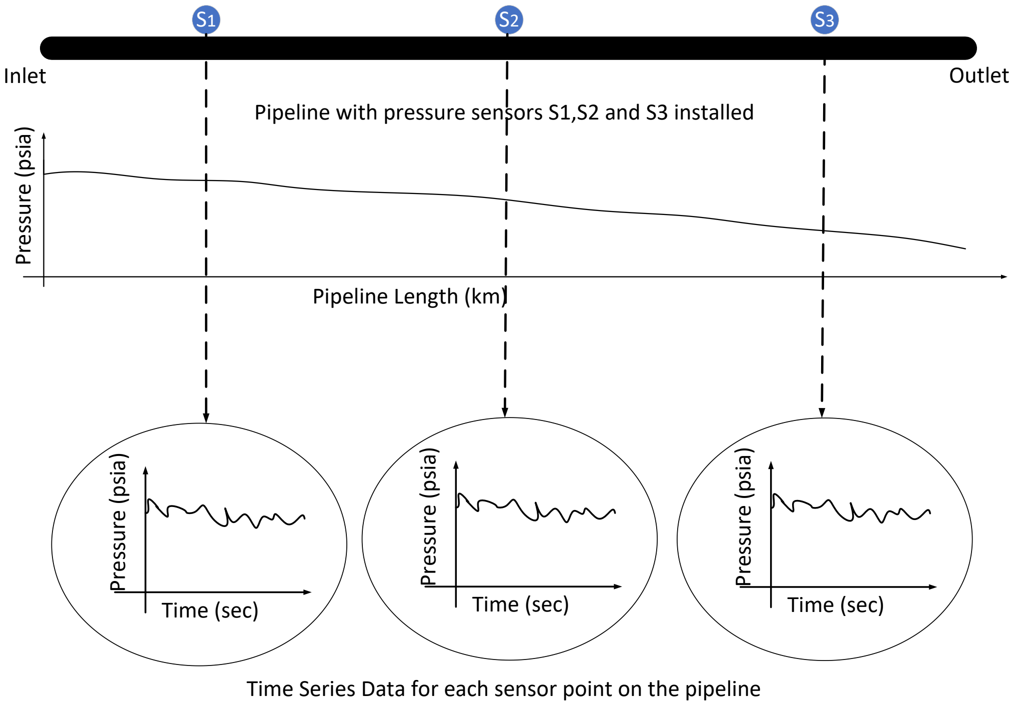


Figure 2.12 Time series data from pipeline sensors

This pipeline pressure profile can be regarded as time series data at each sensor point and the anomalies in the data which infer the presence of leaks or other pipeline control activities can be determined from the sensor data.

Time series data are governed by the following key concepts

1. Trend
2. Seasonality
3. Cyclic
4. Stationary

Time series data can be classified as either Univariate where only one attribute varies with time or Multivariate where two or more attributes vary with time

Pipeline data attributes are multivariate, and they include

1. Pressure
2. Temperature
3. Flowrate

4. Differential pressure

However, the pressure data carries the most information about the fluid flow in the pipeline. The temperature profile has been shown to have very minimal variations while the differential pressure and the flowrate can be derived from the pressure data.

Univariate Anomaly Detection

This is an anomaly detection system that utilizes a univariate dataset. Key advantages of this system include

1. It is easier to model and scale
2. It can learn normal behaviour separately and faster
3. Less data is required for the system to identify anomalies

Multivariate Anomaly Detection

The Multivariate anomaly detection technique utilizes several metrics and combines the signals as relating to one system without separating them. This is an anomaly detection system that utilizes a multivariate time series dataset. Characteristics of this system include the following:

1. It requires a single model for all the attributes for the system to perform anomaly detection
2. It is harder to scale and interpret
3. All metrics need to be homogenous

The system considers all the metrics and generates an output indicating the presence of the anomaly without identifying the metrics responsible for the anomaly. For this research, the pipeline pressure data, and a univariate data stream will be used for the development of the anomaly detection algorithm.

A Hybrid Approach

Hybrid anomaly detection combines the advantages of both the univariate and the multivariate anomaly detection methods to develop a system that is easier to implement and scale while at the same time not generating a large number of false positives. This is achieved by analyzing the metrics independently and identifying the anomalies of each metric and then grouping the anomalies. (Anodot 2017). Table 2.3 shows the key characteristics of the Univariate, Multivariate and Hybrid approaches to anomaly detection (Anodot 2017)

Table 2. 3. Comparison of the different Anomaly detection methods (Anodot 2017)

Univariate Anomaly Detection	Multivariate Anomaly Detection	Hybrid Approach
➢ Learn normal model for each metric ➢ Anomaly detection at the single metric level ➢ Easier to scale to large datasets and many metrics ➢ Causes anomaly storms—can't see the forest from the trees ➢ Easier to model many types of behaviors	➢ Learn a single model for all metrics ➢ Anomaly detection of a complete incident ➢ Hard to scale ➢ Hard to interpret the anomaly ➢ Often requires metric behavior to be homogeneous	➢ Learn normal model for each metric ➢ Combine anomalies to single incidents if metrics are related ➢ Scalable ➢ Make interpretation from groups of anomalies ➢ Can combine multiple types of metric behaviors ➢ Requires additional methods for discovering the relationships

2.5 Anomaly Detection in Time series Data

Anomaly detection is data driven process for identifying unusual patterns in a data stream. The data which is generated from the operation of a system, or the behaviour of users and systems are characterized by a defined pattern. Any deviation from this pattern is termed as an anomaly and the data is classified as an outlier. Anomaly detection is therefore a technique for detecting these patterns with the view to identifying the causes of the unusual behavior of the system or the users. The taxonomy of Outlier detection in time series data is also represented in figure 2.13.

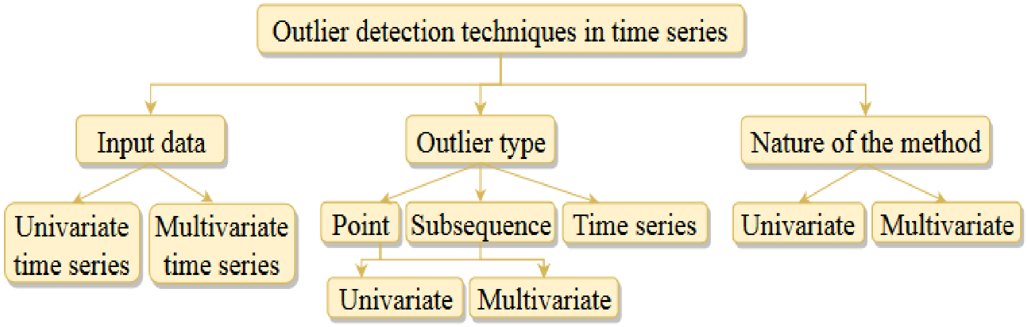


Figure 2.13 Taxonomy of outlier detection techniques (Basu and Meckesheimer 2006).

Anomaly detection relies on historical data, and this assumes that the underlying process which led to the generation of the data is largely constant. This means that the behavior of the system or the user is predictable and can be determined using models and specific algorithms. In cases where the data changes over time, the pattern of these changes is also evident in the data. These data which change over time are defined by some parameters which enable it to conform with specific long-term trends which define the dataset to be time series based. (Kishan et al. 2017).

For real-time systems with the possibility of anomalies, changes in the data make the models developed using those data invalid with the new data. This is because the model is not guaranteed to generalize to the new datasets. Traditional approaches for anomaly detection in time series follow an established process where the model is developed using datasets in an offline manner. This model is now used to detect anomalies in a new unseen dataset. Anomalies occur when these datasets change. The anomaly detection models need to be able to adapt to these changes in real-time. This change which is an indication of a change in normal behaviour is known as concept drift (Saurav et al. 2018). The ability of the models to adapt to these changes is referred to as Concept drift adaptation.

Anomaly Detection for pipeline leak detection relies on the fact that the operation of the pipeline and fluid flow in the pipeline is governed by principles which lead to the predictable operation of the fluid flow process in the pipeline. The principles governing this fluid flow define the normal operation of the pipeline and this can be tracked using the pressure profile of the pipeline. Variations in the pressure profile of the pipeline can be tracked and classified as an anomaly. Processes such as the switching of the pumps and valves and pipeline leaks result in the deviation from the normal operating process resulting in datasets varying from the expected values at specific locations along the pipe. Pipeline leak is the most important anomaly this research seeks to detect. The possibility of several other factors of the pipeline operation leading to anomalies results in the generation of false alarms thereby degrading the performance of the leak detection system. Feature extractions enable the isolation of unwanted anomalies thereby enabling the detection system to identify anomalies associated with pipeline leaks.

These scenarios discussed above are classified as follows

1. Correct Detection: These are cases where the detected abnormalities correspond to the abnormalities in the process

2. False Positives: These are unexpected abnormalities in the data which are not the abnormalities of interest. They can be due to system noise.

3. False Negatives: These are cases where the anomaly detection system fails to classify anomalous data accurately. This can be due to the presence of noise or other data corruption making it difficult for the anomaly detection system to differentiate the anomaly from the normal datasets.

2.5.1 Types of Anomalies (Outliers)

There are different types of outliers in datasets. These are discussed in the following sections.

1. Global (Point Anomaly) Outlier.

This is a case where the outlier data has a value which is far outside the range of the system data value. This is usually classified as a rare event. Nearly all available unsupervised anomaly detection algorithms today are from this type.

2. Contextual Outliers

These are data points whose values do not correspond with what is expected at that point in the same context. Contexts are usually temporal, and the same situation observed at separate times can be not an outlier.

3. Collective outliers

These are a group of data points which deviate from the expected behaviour.

2.5.2 Anomaly detection strategies

Anomalies are usually detected using three key approaches

1. Utilization of rules and thresholds set by domain experts. This is accomplished by setting upper and lower admissible values or taking the standard deviation of all the residuals. This approach is static and not able to adapt to changes in the operational environment of the system being monitored

2. Deviation of key performance indicators from the normal pattern. These indicators are often derived using statistical approaches such as standard deviation, confidence intervals etc. The anomalies are detected when the points fall outside the defined confidence levels or standard deviations. This approach requires domain expertise and would require the use of a window of data for the

analysis to be accurate. This window also known as the moving average assumes that a point is anomalous if it is beyond 3 rolling standard deviations.

3. The Utilization of Machine Learning: This approach enables the development of anomaly detection models which can learn from the data and efficiently detect anomalies in the data. This is accomplished using two key steps.

   (a) Forecast future values from the dataset

   (b) Subtract the predicted values from the actual values to determine the anomaly.

To accurately detect anomalies from the time series data, both the trend and seasonality of the seasonality have to be removed to ensure that the system accurately detects the anomaly. The removal of both seasonality and the trend can be done automatically using any of the following methods

1. Augmented Dickey-fuller test,
2. KPSS test
3. Canova-Hansen test

2.5.3 Deep Learning-based Anomaly Detection System for Pipeline Leaks

Deep learning is a class of machine learning which is used where the datasets are large. The introduction of several hidden layers into an Artificial Neural network converts the network into a Deep Learning network. The large volume of data associated with the pipeline networks creates an opportunity for the application of Deep Learning for the development of the anomaly detection network. The Taxonomy of anomaly detection techniques using machine Learning is shown in figure 2.14.

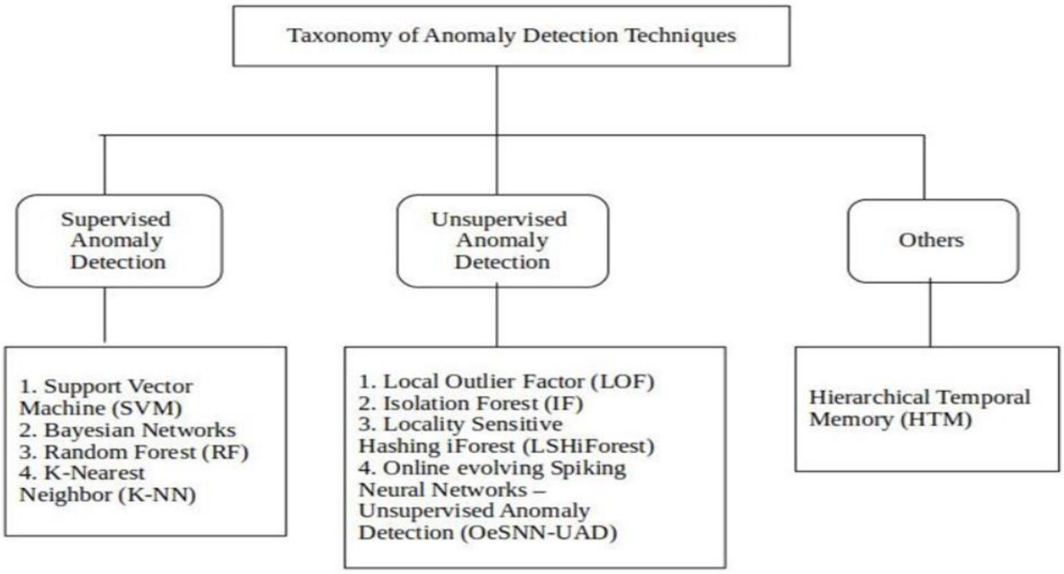


Figure 2.14 Classification of Anomaly Detection techniques (Saranya and Chellammal 2020)

Pipeline leak datasets which comprise of a Univariate stream of time series pressure data can be utilized for the detection of leak occurrences using Deep Learning (Machine learning algorithms). The key machine-learning algorithms which can be utilized for anomaly detection in pipeline data include

1. Supervised learning methods: The key requirement for the utilization of this algorithm is that the datasets must be labelled. The models utilized in this algorithm undertake a classification approach and the result is an output of either normal or abnormal. It is not able to detect new anomalies. It requires a large volume of labelled datasets. They are usually more accurate than unsupervised learning methods and have fewer false positives, but the requirement of a well-defined and labelled data set itemizing all the possible incidents makes this approach very expensive and difficult to implement. It is not easy for organizations to prepare a list of well-defined incidents and everything that could happen especially when these incidents have different anomaly signatures. Typical models used for implementing supervised learning include

a. Neural Networks

b. Decision Trees

c. Support Vector Machine

d. K-Nearest Neighbor

e. Bayesian Networks

2. Semi-Supervised: This method utilizes the subset of the labelled data. All the datasets belong to one class, and it has no definite predictions. It is also known as Novelty Detection. Examples are listed below

a. Neural Networks

b. Clustering

c. Gaussian

d. Tree Based

e. Support Vector Machine

3. Unsupervised learning method: This method uses statistical tests to detect anomalies. The algorithms are fast, and the system learns by itself and determines what is normal over time. The anomaly detection is executed when whenever the current data that is presented deviates from that normal model. The system can detect any type of incident, known or unknown. A key disadvantage of this method is that the performance of the system is defined by the quality of the definition of the normal operation. Since the model has no training sequence. It does not require labelled data and can detect any type of anomaly. Examples of the algorithms used in this learning method include

a. Clustering

b. K-Means

c. Autoencoders

4. Hybrid method: This method utilizes a mix of both labelled and unlabeled data. The supervised and unsupervised methods are combined in this approach where the supervised approach is used to detect known anomalies while the unsupervised approach is used to detect unknown cases.

Modern ML techniques for anomaly detection include

1. Transfer learning
2. Zero-Shot Learning
3. Ensemble learning
4. Reinforcement Learning

Table 2.4 shows the pros and cons of some outlier detection methods

Table 2.4 Summary of outlier detection methods (Iqmal et al. 2020)

Outlier Detection Methods	Descriptions	
Mathematical Function		
Z-score/Extreme Value Analysis	• Discover distribution of the data where mean is zero and standard deviation is equal to one (i.e. normal distribution) • Data points which are away from zero will be treated as the outliers.	
	Pros	Cons
	• A very effective method with a Gaussian distribution • Easy implementation using pandas & scipy.stats libraries	• Convenient to use in a small to medium dataset • Not recommended when distribution is not be assumed parametric
Interquartile Range (IQR)	• A good statistic for summarizing a non-Gaussian distribution sample of data	
Visualization Tools		
Histogram	• To visualize bin that deviates hugely from the rest of the data	
Box Plot	• Graphical method to visualize IQR • Can easily spot the outliers outside the whisker box	
Other Method		
K-Means Clustering	• An unsupervised machine learning that using the Euclidean distance internally to classify different clusters	
Principal Component Analysis	• A linear regression model to compute p-values to test for outliers	
DBscan	Pros	Cons
	• Super effective method when the distribution cannot be assumed • Work well for searching outliers in multidimensional data • Easy implementation with Sci-kit learn • Intuitive result visualization	• Value need to be scaled • Challenge in selecting optimal parameters • Need to be re-calibrated for each time with a new batch as it is unsupervised model
Isolation Forest	• A tree-based anomaly detection algorithm.	
	Pros	Cons
	• No need of scaling the values • Effective method when value distributions cannot be assumed • Fairly robust & easy to optimize as it has few parameters • Easy implementation with Sci-kit learn	• Visualizing results is complicated • Training time can be very long and computationally expensive

Anomaly Detection Using Unsupervised Learning Methods

The majority of the pipeline profile datasets are unlabeled time series data as such the unsupervised anomaly detection algorithms are used in the development of the anomaly detection models. Characteristics of some of the unsupervised anomaly detection algorithms are discussed in the following sections. (Chalapathy and Chawla 2019)

1. Clustering Algorithms

The clustering algorithms are insensitive to time, so the information is not sequential. While this is not in conformity with the time series data models, time classes such as weekday/weekend can be created from the dataset and the anomalous points in these classes detected. Examples of algorithms utilized under this approach include

(a) Isolation Forest

(b) One-Class Support Vector Machine

(c) Histogram-based Outlier Score

2. Autoencoders

These are more refined algorithms when compared with the clustering algorithm. The work by compressing the time series metrics into lower dimension latent space and then reconstructing the data again. Examples of these algorithms include

(a) LSTM-based Autoencoders

(b) Multichannel CNN encoders /LSTM Decoder

3. Time Series Forecast :

This is the best and most widely used anomaly detection approach. In this method, a forecast of the future values is made with time stamps and these values are compared with the real-time values. The anomalies are detected from these comparisons. Examples of algorithms used in this method include

(a) Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU)-based Forecasters

(b) 1d CNN Forecasters

(c) Specific Time series forecasters such as SARIMA or Holt-Winter Exponential Smoothening

(d) Regressors such as XGBoost, Light GBM etc.

Hierarchical Temporal Memory

Deep learning networks are faced with some limitations which limit the quality of the results they can generate and the areas of application they can be deployed in. One of these challenges is the fact that they require historical data and as such are not able to work on real-time anomaly detection systems. Some other limitations of the deep learning algorithms include the following:

1. Requires thousands and in some cases millions of data samples for training
2. Cannot easily adapt to continuously changing (streaming data)
3. There are susceptible to noise and can be easily fooled.

These limitations led to the development of the Hierarchical Temporal Memory (HTM) approach. The HTM is a theoretical framework for both biological and generalized machine intelligence. It is based on the latest understanding of the neocortex of the brain. (Khan et.al., 2021;Mountcastle 1998; Billaudelle and Ahmad 2015)

Key advantages of this approach include the following:

1. It requires only a few hundred samples to be able to learn
2. It learns unsupervised and can adapt easily to changing data
3. It is immune to up to 40% of noise in the data.

The HTM works with streaming data which requires that the data changes over time. It can handle the following tasks

1. Prediction
2. Anomaly Detection and
3. Classification

The block diagram of the HTM approach for anomaly detection is shown in figure 2.15

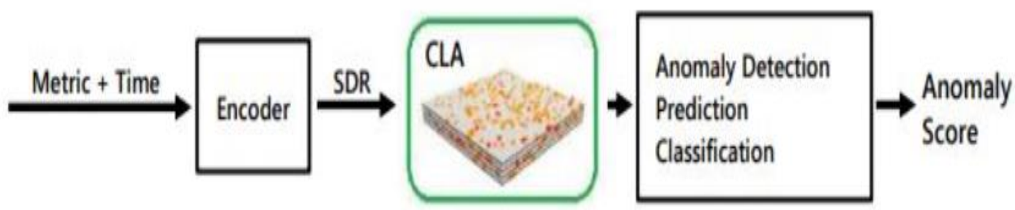


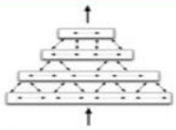
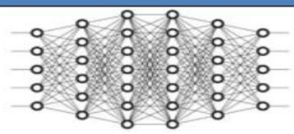
Figure 2.15 Block diagram for Applications using HTM High-Order Inference (Ahmad and Hawkins 2017)

Hierarchical Temporal Memory (HTM) is a learning system capable of continuously learning from the environment. It is also referred to as an online learning system. (Mountcastle 1998; Billaudelle and Ahmad 2015). It operates by detecting anomalies in real-time from streaming data. The input data for HTM functionality is either numerical or categorical. The HTM system merges both data types into an input data stream and is converted to a sparse distributed representation using encoders. The HTM system then calculates an anomaly score for every new pattern that it receives from the input data stream (Görnitz et al. 2013; Ahmad and Hawkins 2017; Hawkins et al. 2017). If the received data values are the same as the predicted values, the anomaly score is computed to be zero but if they are different, then the anomaly score is computed to be one. The SDR of the input data stream determines the similarity. The “similarity” between the actual received values and the predicted values is the base of The HTM score. The larger the overlap between actual and predicted input patterns results in a smaller anomaly score. (Khan et al. 2021). This approach opens the way to the development of truly intelligent systems. The suitability of the HTM system for online learning is based on the structure of the HTM design. They are designed to work on a continuous temporal stream of data which changes in time. This differs from most artificial neural network techniques that are designed to require massive static datasets.

The learning algorithm of the HTM systems is known as Hebbian Learning and it forms the basis of learning mechanisms in HTM systems. Hebbian Learning is a phenomenon where repeated activity between neurons strengthens their connections and vice versa. The synapses on every dendritic segment of every HTM neuron are updated with each new change in the data signal. Learning in the HTM systems is thus completely local as they occur at the level of the synapses. (Brody 2018)

A comparison between the HTM and the Deep Learning algorithms is shown in Table 2.5

Table 2.5 Comparison between HTM and Deep Learning Algorithms(Brody 2018)

Attribute	HTM (Hierarchical Temporal Memory)	Deep Learning
		
Premise	Biological	Mathematical
Learning Mechanism	Hebbian Learning	Back Propagation
Learning type	Unsupervised	Supervised
Learning batches	Online learning	Batch wise learning
Neuron cell state	Active/Inactive/Predictive	Active/Inactive
Batch size need to learn	Very small data is sufficient	Required huge data volume

2.5.4 Real-Time and Batch Anomaly Detection

Anomaly Detection can also be implemented both in the batch mode and in the real-time mode.

Batch Mode Anomaly Detection

The characteristics of the batch mode anomaly detection system include the following:

1. Time is not a critical factor for the detection of the anomaly
2. The system requires a large volume of data for the detection of the anomaly
3. The algorithm can iterate over the data multiple time
4. The process is computationally expensive and scales very poorly.

Real-Time Anomaly Detection

The characteristics of the real-time anomaly detection system include the following

1. The system can learn continuously from streaming data without having to store the entire data stream.
2. The system is not manually supervised
3. The anomaly detection model evolves as the behavior of the system changes

Considerations for Real-time Anomaly Detection

The key considerations for the deployment of real-time anomaly detection systems include

1. Timelines: In what time interval does the business need to know of the anomaly
2. Scale: How large is the data required to be processed to determine the anomaly
3. Rate of change of data. Does the data change quickly or is it static?
4. Conciseness: Will there be a need for multiple metrics to produce an answer from the anomaly detection system?
5. Definition of Incidence: are the anomalies well known are they labeled?

While online Anomaly detections have a very high response time, they are also prone to the generation of False Positives.

For algorithms to qualify to be online or real-time, they are required to meet the following conditions (Brody 2018)

1. Predictions must be made online, i.e., the algorithm must identify state x_t as normal or anomalous before receiving the subsequent x_{t+1}.
2. The algorithm must learn continuously without a requirement to store the entire stream.
3. The algorithm must run in an unsupervised, automated fashion without any data labels or manual parameter tweaking.
4. Algorithms must adapt to dynamic environments and concept drift, as the underlying statistics of the data stream are often non-stationary.
5. Algorithms should identify anomalies as early as possible.
6. Algorithms should minimize false positives and false negatives.

An algorithm that possesses these properties can rightfully be called a streaming Anomaly Detection algorithm.

2.6 Performance Metrics for Anomaly Detection

The detection of anomalies in a data stream begins with the accurate modelling of the normal behavior of the system. The anomaly detection model must be able to accurately model the normal behavior of the system to be able to detect anomalies when they occur. This approach is implemented in a 3-step strategy listed below

1. Model the normal behaviour using forecasting and statistical models
2. Generate statistical parameters to represent the operational envelope of the normal behaviour
3. Apply the statistical test to each data point and flag as anomalous any point that deviates outside the set operational envelop

The most predominant statistical test utilized in modelling the normal behaviour of the system is the normal distribution. This is represented by the expression in equation 2.12 (Krithikadatta 2014)

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad 2.12$$

Where:

σ =The standard deviation,

It is sensitive to the presence of anomalies as such it is used in determining the anomaly. The test for the anomaly is presented below

1. Determine the forecast value of the data at a particular point
2. Compare the actual data at that point with the forecasted data at that point
3. If the difference between the actual data and the forecasted data is greater than 3 times the standard deviation at that point, then the data is classified as an anomaly

Figure 2.16 shows the standard deviation chart.

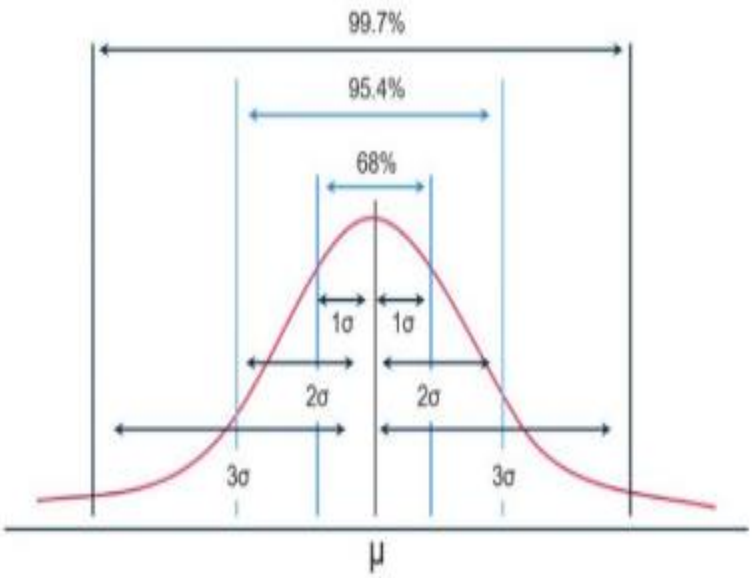


Figure 2.16 Normal distribution curves showing the anomaly threshold points. (Krithikadatta 2014)

Scoring Anomalies

The Z -score is a statistical parameter used in determining the presence of an anomaly. It is applied to 1D data and calculated for each data point using the expression in equation 2.13.

$$Z_i = \frac{x_i - \mu}{\sigma} \qquad 2.13$$

The z-score measures how far a point is from the mean value of that data point. Large values of the z-score indicate the presence of an anomaly. However, studies have shown that the sensitivity of both the mean and the standard deviation makes the z-score metric an unreliable parameter for the determination of the presence of anomalies. This led to the use of the modified z-score which relies on medians. The expression for the modified z-score is given in equation 2.14

$$Z_{mi} = \frac{x_i - \hat{X}}{MAD} \qquad 2.14$$

Where $\hat{X} = Median\ of\ X$, $(MAD = median(|x_i - \hat{X}|)$, $Zmi = modified\ z - score$

MAD = **M**edian **A**bsolute **D**eviation from the median

Large absolute values of the modified z score indicate the presence of an anomaly.

Raw Anomaly Score Calculation. The Raw anomaly score is required for the computation of the difference between the actual and the predicted data values of a given point at any specific point in time. It is computed from the intersection between the predicted and actual sparse vectors in HTM. (Khan. et al 2021)

Output of Anomaly Detection Algorithms

The output of anomaly detection systems are determined by the machine learning algorithm used in implementing the anomaly detection system. For supervised anomaly detection, the output is a classification-based output which is defined by the label. The label indicates if the data instance is an anomaly or not. For unsupervised or semi-supervised anomaly detection, the output is presented as a score or a confidence value which indicates the degree of abnormality. Scores are preferred as they can be easily ranked and only the top anomalies are flagged. (Goldstein et al.2016).

Performance metrics of Anomaly Detection Systems

The performance metric of the anomaly detection system depends on the type of machine learning algorithm used. For supervised learning, the metrics used for evaluating the anomaly detection algorithm are shown in Table 2.6

Table 2.6 Anomaly Detection Metrics

Confusion Matrix	**Actual Normal Data (n_n)**	**Actual Anomalous Data (n_a)**
Predicted Non-anomalies	TN	FN
Predicted Anomalies	**FP**	**TP**

TN = True Negative is the number of correctly predicted non-anomalies.

TP = True Positive is the number of correctly predicted anomalies.

FN = False Negative is the number of actual anomalies which are predicted as non-anomalies.

FP = False Positive is the number of non-anomalies which are predicted as anomalies.

True Negative Rate (**specificity**): the ratio between the number of correctly detected normal data (TN) and the total number of normal (TN+FP) data

Precision: the ratio between the number of normal data that are misclassified (FP) as anomalies and the total number of data records from normal class (TN+FP)

recall (**sensitivity**): the ratio between the number of correctly detected anomalies (TP) and the total number of anomalies (TP+FN)

Accuracy rate (ACC), is a traditional metric used to evaluate the classifier performance in the community of data mining and machine learning it represents the percentage of right prediction from the entire datasets. (Hajian-Tilaki 2013).

$$Accuracy\ Rate\ (ACC) = \frac{(TP+TN)}{(TN+FP+FN+TP)} \qquad 2.15$$

True Positive Rate (TPR) represents the percentage of anomalies that are correctly detected, i.e., the ratio between the number of correctly detected anomalies and the total number of anomalies.

$$True\ Positive\ Detection\ Rate\ (TPR) = \frac{TP}{(TP+\ FN)} \qquad 2.16$$

False Alarm Rate (FAR) or **False Positive Rate (FPR**), represents the percentage of normal data that are incorrectly considered as anomalies, i.e., the ratio between the number of normal data detected as anomalies and the total number of normal data.

$$False\ Alarm\ Detection\ Detection\ Rate\ (FAR) = \frac{FP}{(FP+\ TN)} \qquad 2.17$$

Receiver Operating Characteristics (ROC) curves are two dimensional plots in which the true positive (TP) rate is plotted on the Y axis and the false positive (FP) rate on the X axis

The Area under the curve (AUC) of a classifier is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance. The AUC-ROC curve is a performance measurement for the classification problems at various threshold settings. ROC is a probability curve and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. The larger the AUC, the more robust the classifier is. (Hajian-Tilaki 2013). In Figure 2.17, Curve B represents a more robust classifier while the entire square is Ideal AUC.

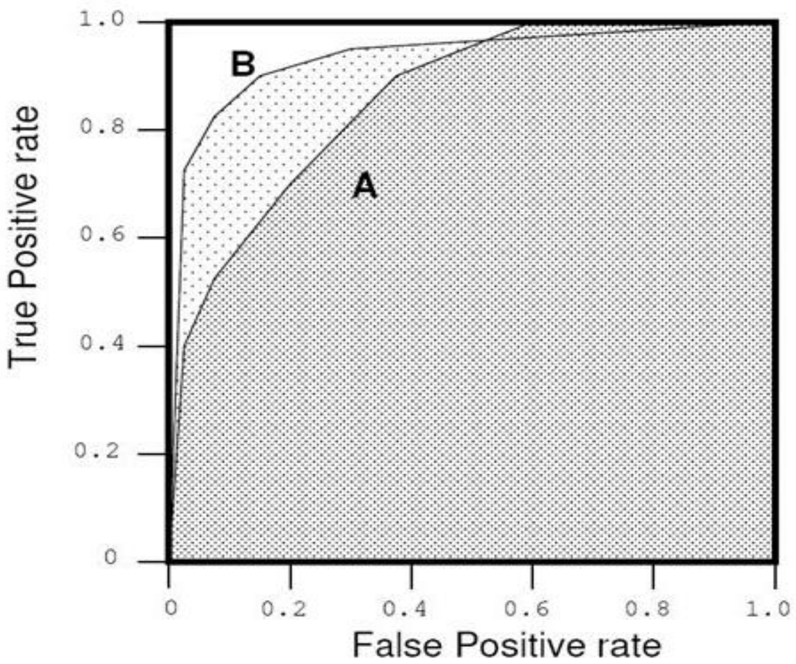


Figure 2.17 Area Under the curve

2.7 Unsupervised learning algorithm for pipeline leak detection

Pipeline data can be classified as time series data. These data vary with time and are usually unlabeled. A review of the pipeline data shows that of all the metrics, the pressure is the most representative of the pipeline flow status. The pipeline pressure data can thus be used to detect the operating conditions of the pipeline as well as detect anomalies in the pipe flow. Variations in the pipeline pressure profile which result in outliers range from switching activities of either the pumps or the well to the presence of leaks or pipeline ruptures. The difference between the outliers and the anomalies is that while the outlier is a variation in the pressure profile, the anomaly is a variation of interest. Anomaly detection in the pipeline data is therefore aimed at detecting pressure variations occasioned by the presence of a leak in the pipeline. The leak detection algorithm will be categorized as an anomaly detection algorithm and will utilize unsupervised learning algorithms due to the lack of labels on the pressure data from the pipeline.

2.7.1 Data Analytics Applications for Leak Detection (Selected Case studies)

Several researchers have utilized Machine learning tools for the development of leak detection systems, and they all have reported different success levels. Artificial Neural Networks (ANN), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF) and Gradient Boosting, have been utilized by researchers with varying levels of sensitivity, accuracy, and reliability. The key findings of the research suggests that data analytics and artificial intelligence can be utilized with the RTTM to improve the leak detection results (Akinsete and Oshingbesan 2019; Kang et al. 2018). The quality of the data acquired from the sensors is very critical to the quality of the output from the Machine Learning models used for the development of the leak detection algorithms ((Romano et al. 2011).

(Caputo and Pelagagge 2003; Sivapragasam et al. 2007) proposed the application of an Artificial Neural Network (ANN) based approach for the leak detection system. Their approach used pressure and flow to infer the leak location and severity, through an ANN trained on a dataset generated by a mathematical model of the network and the hydraulic simulation software EPANET. The Support Vector Machine approach was proposed by (Candelieria et al. 2014; Mashford et al. 2012).They combined the EPANET-based leakage simulation software and SVM. The SVM model was trained on a dataset of leaks simulated on the junctions of the water distribution networks. (Most approaches simulate leaks on pipes). The trained SVM classifier was able to identify the leaky junction(s) by utilizing only the pressure and flow values (Candelieri et al. 2014). Hidden Markov Model-based agents (Nasir et al. 2010), Genetic Programming (Lijuan et al. 2012).), and Bayesian approaches ((Poulakis et al. 2003; Xia et al. 2006) are some other prominent machine learning algorithms being utilized for the development of leak detection systems. Table 2.7 presents a summary of key findings from the use of different data mining strategies and models for leak detection.

Table 2.7 A summary of Data Analytic models for leak detection.(Idachaba and Minou 2021)

Method	Authors	Key findings
Artificial Neural Network	(Caputo and Pelagagge 2003; Mounce et al. 2010; Salam et al. 2014; Zhang et al. 2016; Romano et al. 2011;Mounce et al. 2007; Huang et al. 2018; Bohorquez et al. 2020)	ANNs rely too much on training samples. The ANN prediction accuracy is poor if the size of the training samples is small. To achieve adequate accuracy, sufficient training samples will be needed. The ANN training time increases as the size of the training samples increases. ANN needs to be retrained when the physical conditions of a water system change; this is quite common in developing countries,
Support Vector Machines	(De Silva et al. 2011)	They found that the predicted leak location was within 500m of the actual leak location in all cases for a network that could fit into a 1000 by 1100m square box. The smallest leak registered by EPANET to generate a pressure difference was a leakage of 90l/hour. A new data set was created to which the SVM was trained. A testing accuracy of 35% was found.
Bayesian Probabilistic Framework	(Poulakis et al. 2003;Zhou et al. 2011;Costanzo et al. 2014;Romano et al. 2010)	They found that when the model measurements had an uncertainty 5% the model couldn’t determine the actual leak location.

2.8 PIPESIM Simulation Software

The Schlumberger PIPESIM is a steady-state multiphase flow simulator capable of offering complex production and injection network analysis. It is a computational fluid dynamics software (CFD) that enables the accurate replication of petroleum production systems. By modelling the entire production or injection system as a network, the interdependency of wells and surface equipment can be accounted for, and the deliverability of the system can be determined. (Schlumberger 2022).

PIPESIM offers comprehensive steady-state flow assurance workflows for front-end system design and production operations for the oil and gas industry. The flow assurance capabilities of the simulator enable the design of safe and effective fluid transport—from sizing of facilities, pipelines, and lift systems, to ensuring effective liquids and solids management, to well and pipeline integrity. It also enables the accurate prediction of the behaviour of different fluids and pipeline networks and provides a very high level of flow assurance metrics to the organization by simulating how these fluids will perform during transport and storage. PIPESIM can accomplish all these features by employing a wide range of industry standard multiphase current correlations as well as advanced three-phase mechanistic models. These models enable the software to be able to calculate the flow structure, fluid latency, material motion characteristics, and pressure reduction of all stations along the production path. PIPESIM provides two choices for Liquid Characterization Modeling. The first option is the use of the industry standard black oil concepts while the other is a range of hybrid models covering a wide range of fluids.

PIPESIM has other features which provide the user with the ability to accurately simulate a wide range of scenarios and conditions. One of the critical features is the GIS capability. With PIPESIM, actual pipeline networks can be reproduced using the topological coordinates and the GIS maps of the pipeline right of way. This will enable the simulation of the actual pipeline network with the environmental and topological conditions factored into the simulations. This leads to the generation of results that correlate very well with the actual data generated from the field. Production optimization is another feature of the PIPESIM software. It provides a complete set of workflows which help to achieve optimal production of oil and gas. PIPESIM software can also be integrated with other software production platforms such as the ECLIPSE industry-reference reservoir simulator, Aspen HYSYS, Honeywell UniSim, and KBC Petro-SIM, as well as real-time data for online optimization, Petrel E&P and Avocet to provide a single solution and a complete simulation from tanks to production.

PIPESIM also includes a fully documented application programming interface (API) called Python Toolkit. This API facilitates communication with PIPESIM models directly without opening the User Interface (UI). It also streamlines several functions such as building modelling from scratch, updating existing models, running simulations and getting results back to Excel to any visualization dashboard using Python language.

The numerous simulation and experimentation capabilities of the software eliminate the need for expensive and time-consuming experiments and scenario simulations that are needed to improve the production process in many oil and gas operations. (Alpandi et al. 2021; Schlumberger 2022; Li et al. 2013; Nsofor et al. 2020; Ubani et al. 2018; Prosper et al. 2019).

2.9 One-Class Classifiers

Pipeline leak datasets are predominantly comprised of pressure profiles during normal operations and very minimal instances of leak-driven pressure variations. These leak-driven pressure variations which are referred to as outliers and used for the determination of leak occurrences in the pipeline are much fewer when compared with the no-leak case. With this characteristic, having a labeled dataset with the leak points labelled will not be suitable for machine learning applications due to the class imbalance problem. Thus, the determination of the suitable machine learning model to be deployed in the determination of the leak detection system will rely on the characteristics and nature of the pipeline dataset.

Characteristics of pipeline leak datasets include

1. Majority of the data is in one class (No leak case)
2. The leaked data is very minimal as it occurs only when there is a leak
3. The data is mostly unlabeled
4. The data is subject to class imbalance challenge

With these characteristics, the determination of leaks can be implemented as an anomaly detection problem and the most suitable machine learning model for the detection based on the characteristics of the data is the One class classifier model.

2.9.1 One class classifier Model:

The goal of an anomaly detection system is to detect the presence of anomalous and defective patterns that are different from the expected normal data stream. These applications range from use cases in manufacturing defect detection (Carrera et al. 2015; Bergmann et al. 2019), medical image analysis (Schlegi et al. 2017), and video surveillance (Liu et al. 2018). One major challenge faced in the development of anomaly detection systems is the difficulty in obtaining a large amount of anomalous data. The second challenge with anomaly detection is the imbalance between the normal and the anomalous datasets. This is known as the class imbalance problem. When the number of samples in one class is a lot greater than the number of samples in the other classes, the question of class imbalance arises. (Rekha et al. 2021). Traditional machine learning algorithms assume that the number of objects in the classes of interest in a dataset are equal. This assumption is not the case in real-world applications as the distribution of examples is skewed since representatives of some classes appear much more frequently. This distribution presents a challenge for the learning algorithms as their outputs will be biased towards the majority class. (Krawczyk 2016). Due to limited access to anomalous data, and the challenge of class imbalance, constructing an anomaly detector is often conducted under semi-supervised or one-class classification settings using normal data only. One class datasets comprise datasets with only one class of the classification instance. The objective of the classification model is to determine any data point that is not within the normal datasets and classify such data points as anomalies. To accomplish this, the dataset is expected to be a one class dataset. There are three approaches to handling class imbalance problems in datasets. These approaches are

1. Data-level methods: These are methods that modify the collection of samples to balance the distributions and/or remove the difficult samples
2. Algorithm Levels methods: These are methods that directly modify existing learning algorithms to alleviate the bias towards majority objects and adapt them to mining data with skewed distributions.
3. Hybrid methods: These are methods that combine the advantages of both the Data level methods and the Algorithm Level methods.

Chapter 3

Materials and Methods

In this Chapter, the methodology utilized for the research is presented. The methodologies are designed to address the research questions identified from the literature search.

3.0 Materials and Methods

The research utilized the simulations based approach and the PIPESIM simulation software was used for the simulations of the pipeline network. The software was selected due to the correlation of its simulation results with field data and the GIS functionality which enables the modeling of actual operator pipeline networks

3.1 Research Question 1

How can pipeline leak and no-leak datasets for specific pipeline networks be generated using simulation software

3.1.1 Research Question 1: Methodology

1. Simulation of a 20km straight pipeline system using PIPESIM to generate the pipeline pressure profile and determine the model of the pressure profile.

2. Generation of the time series pressure data from selected sensor location on the pipeline using the RAND function in python
3. Simulation a leak on a 20 km straight pipeline by connecting a choke and sink to the mid-point to represent the leak point.
4. Generation of time series leak datasets for the pipeline using the RAND function from python.

3.1.2 Experimental Processes for Research Question 1

(a) Generation of Pipeline Pressure Profile

A 20km horizontal pipeline network was implemented using PIPESIM with the following parameters.

Inlet:

The inlet pressure is selected to be 1000psia and the black Oil fluid is used in the simulation. The Inlet settings are shown in figure 3.1

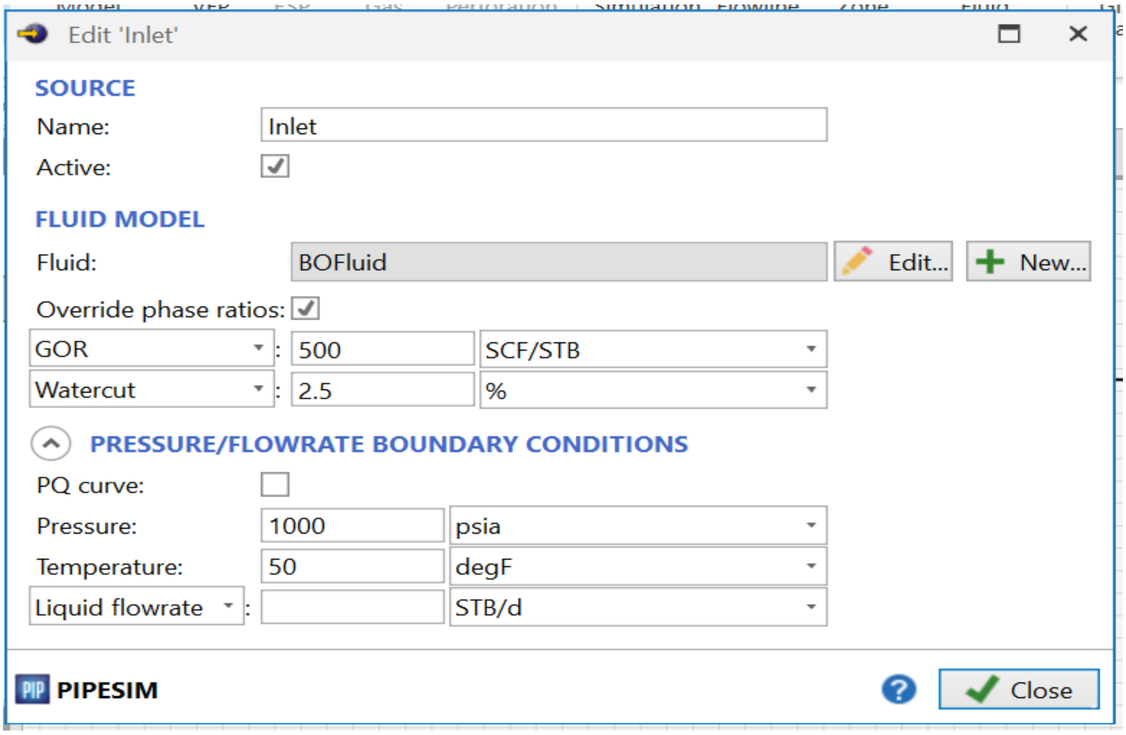


Figure 3.1 Inlet parameters

The Outlet.

The pressure at the outlet is set to 400psia. The other parameters are shown in figure 3.2

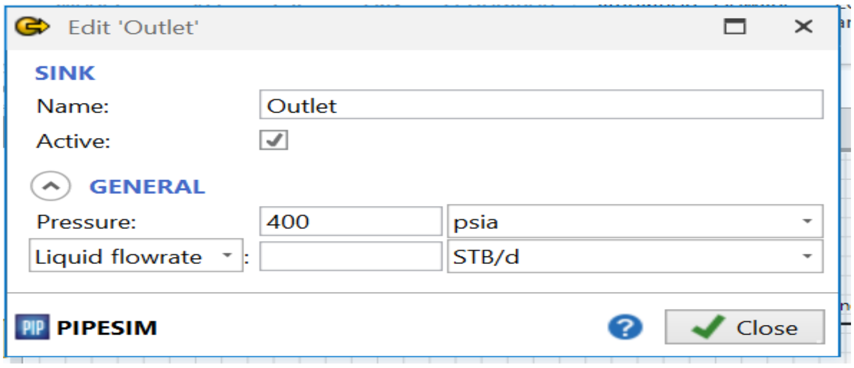


Figure 3.2 Outlet parameters

Pipeline

A 20km pipeline is selected for this experiment. The parameters of the pipeline are shown in Figure 3.3.

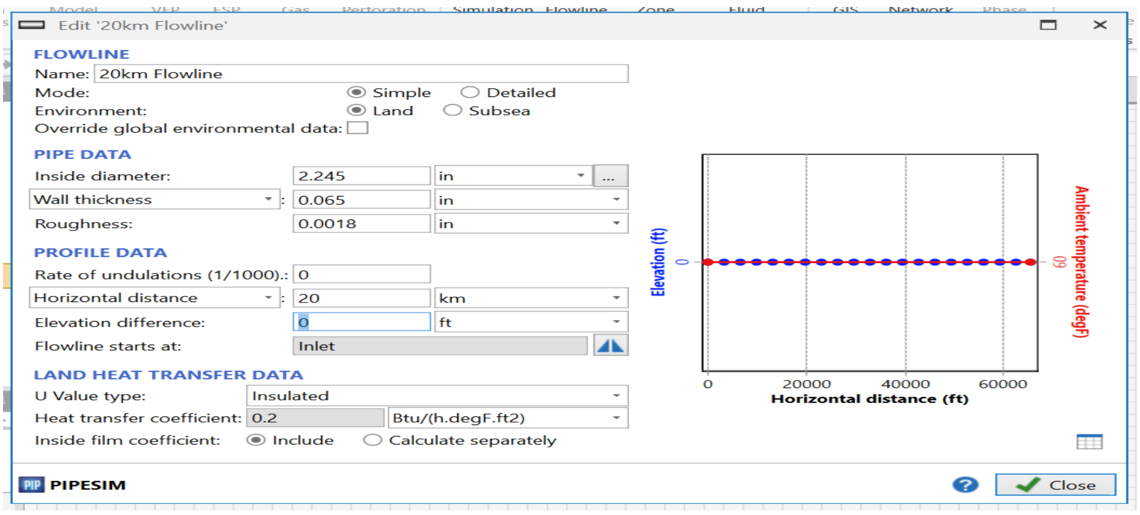


Figure 3.3 Pipeline Parameters

The complete experimental setup for the pipeline pressure profile is shown in Figure 3.4

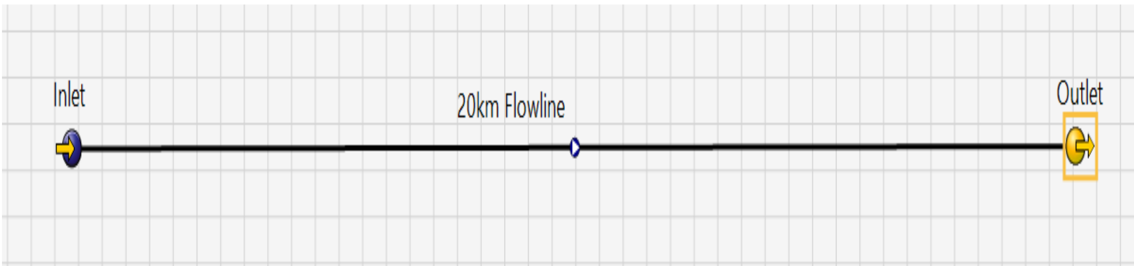


Figure 3.4 Complete experimental setup for 20km pipeline

The network simulations are run multiple time and the datasets for the pipeline pressure profile is generated.

(b) Generation of Time Series Data from pipeline pressure profile

The time series datasets for the pipelines consist of time-varying data from a point on the pipeline. These data are usually acquired using sensors installed at specified locations. This data is a continuous stream of data that varies with time around an average point.

The steps for the generation of the time series data sets from the pipeline are as follows

(i) Determine the equation for the pipeline profile. Generate the pressure values for the different portions of the line using the pressure profile model.

(ii) Select a point on the pipeline for which the sensor is to be installed

(iii) Read the pressure profile data from the table or use the model to determine the pressure value at that point

(iv) Use the RAND function in python to generate a time series data around the pressure profile value at that location.

Generating the first 500 variations of the pressure data at the selected sensor location (J) which is at the middle of the pipeline (10km), the RAND function (Random number generator) of python with a peak-to-peak pressure variation of 10 psi around each point on the pipeline is used. The sensor data is shown in Table 3.1.

Table 3. 1. Sensor Data

Sensor Data	Parameter
Sensor Location	10km
Pressure at Sensor Location	764psia (rounded to nearest integer)
Inlet pressure	1000psia
Pressure fluctuation value	5psia
Pressure range at sensor location	759 -769psi

The python script is shown below

```
import random
a=[random.randint(759, 769) for I in range (0,500)]
print (a)
```

(c) Simulation of a single point leak on the horizontal pipeline

Generating the pipeline leak dataset with the following parameters

Leak location = 10km

Pressure at leak point read from the pressure profile

A 20km horizontal pipeline network with a leak location at the midpoint was implemented using PIPESIM. The experimental process is shown below.

Inlet:

The inlet pressure is selected to be 1000psia and the black Oil fluid is used in the simulation. The Inlet setting are shown in the Figure 3.5.

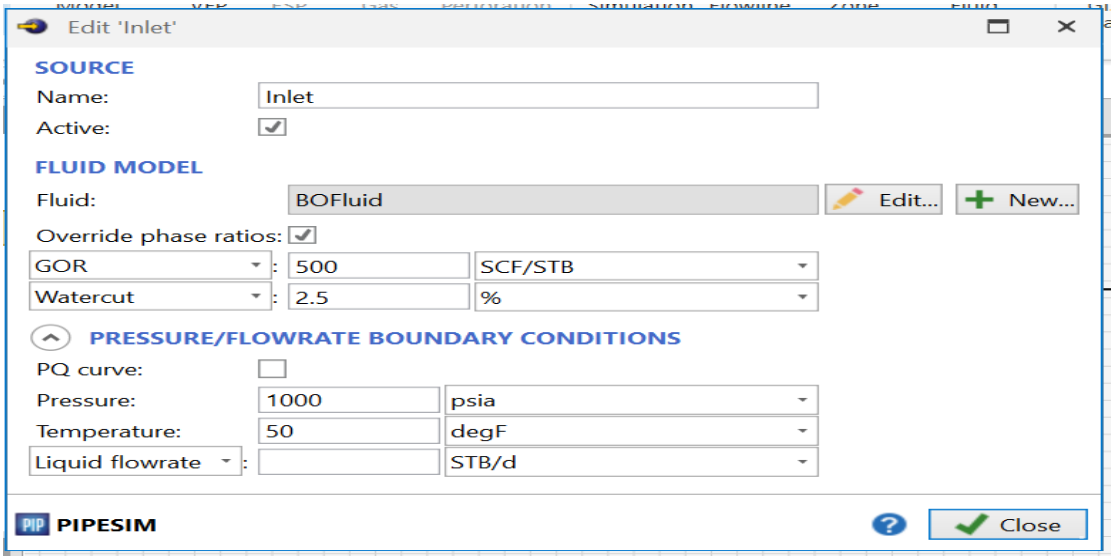


Figure 3.5 Inlet parameters

The Outlet.

The pressure at the outlet is set to 400psia. The other parameters are shown in the Figure 3.6.

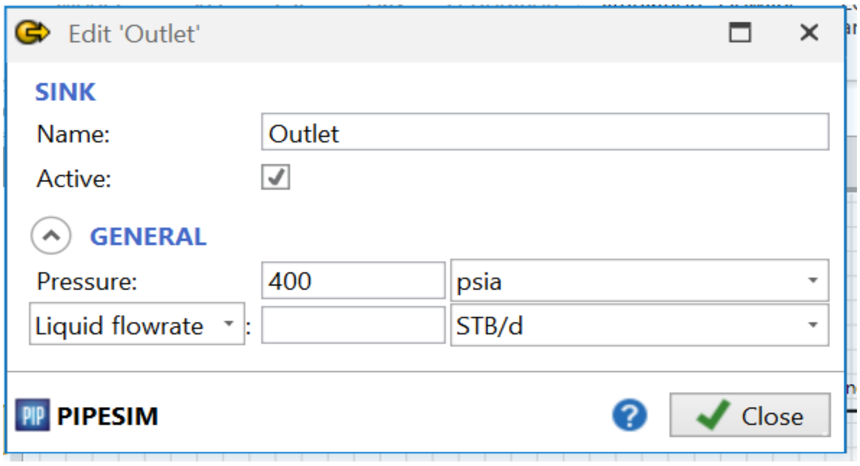


Figure 3.6 Outlet parameters

Leak Valve

The leak valve selection is shown in Figure 3.7.

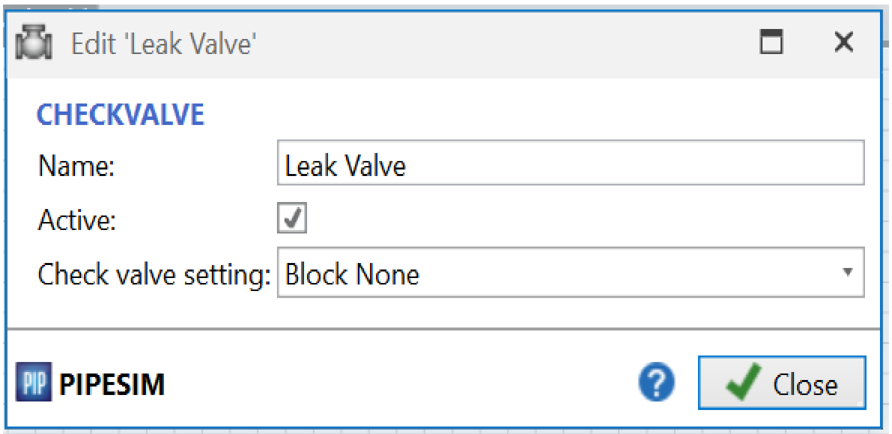


Figure 3.7. Leak valve selection

Pipeline

A 20km pipeline is selected for this experiment. The parameters of the pipeline are shown in Figure 3.8.

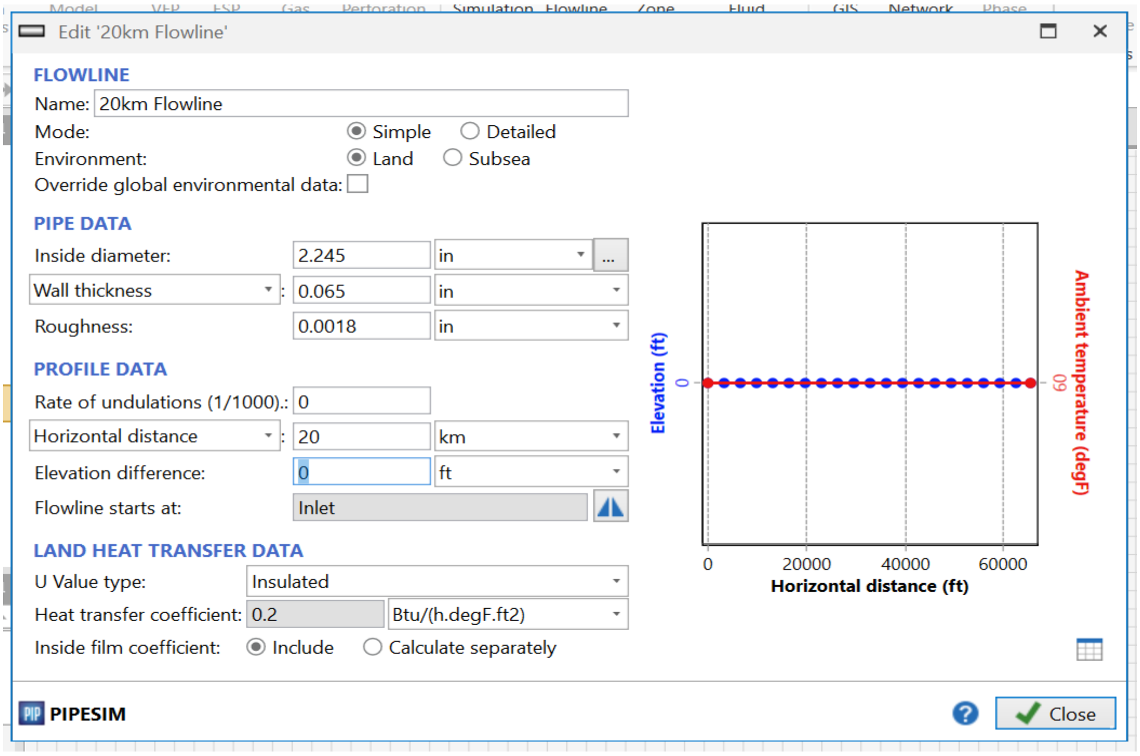


Figure 3.8 Pipeline Parameters

The complete experimental setup for the pipeline pressure profile is shown in Figure 3.9

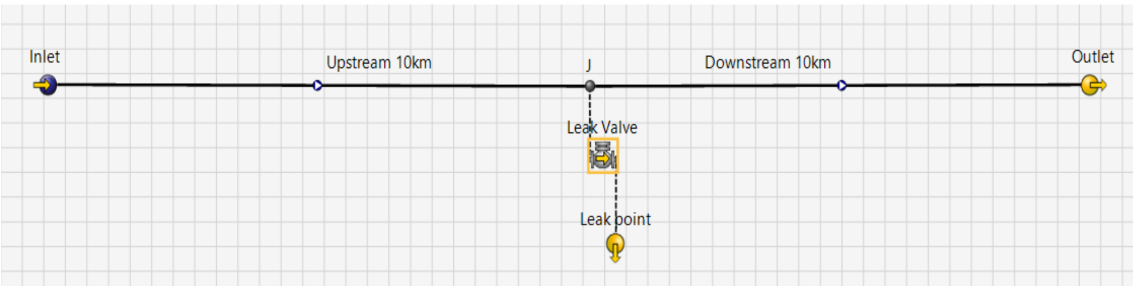


Figure 3.9 Complete experimental setup for 20km pipeline with leak point at 10km

(d) Generation of Pipeline Leak Datasets

The occurrence of a leak divides the pipeline system into three parts: exact leak location, upstream, and downstream. (Araújo et al. 2014; Edrisi and Kam 2013; Sousa et al. 2016). The generation of time series pipeline leak datasets for the leak solution utilizes the pressure data extracted from the pipeline profile for the selected sensor location and uses the following protocol.

1. The pipeline is broken down into 3 sections namely the upstream (pressure value before the leak) pressure value at leak point and the downstream (which is the pressure value after the leak).
2. The selected pressure values for the 3 sections listed are used as seed value for the RAND functions to generate the time stream data for each of the sections
3. The generated time series data for each of the sections are combined in series to produce a continuous leak time series dataset for the pipeline

```
#Import libraries
import matplotlib.pyplot as plt
import numpy as np
import random
#Generate upstream datasets (Pressure value before leak)
# Sensor Location is at mid-point  (10km) the No leak value
# at that point is given to be 764 psi.
# introducing a peak-to-peak variation of 10psi on the pressure value at this point
#to account for the variations in the pressure due to the pipe internal surface area
# and the flowrate at the inlet  import random. Pipeline pressure variation will be
#from 759 to 769psi for a time frame of 500  readings
a=[random.randint(759, 769) for i in range (0,500)]
# Generate Leak point datasets (Pressure value at leak point)
# The pressure value at the sensor due to a leak which occurred 2km from the inlet
# The sensor pressure value due to the leak is given to be 686 psi. Adding a 10psi
# peak to peak  variation for a short duration of 5 readings.
b=[random.randint(681, 691) for i in range (600,605)]
# Generate downstream datasets (Pressure value after the leak)
# Sensor Location is at mid-point  (10km) the pressure value after the leak is such
 # that it rises  from the leak point but does not get to the upstream value. Selecting
#a value of 745psi with  a peak-to-peak variation of 10psi on the pressure value at
# this point with a time frame of 500 readings
c=[random.randint(740, 750) for i in range (605,1105)]
# Combine all the three sections of the pipeline datasets
```

```
d = a+b+c
print (d)
plt.plot(d)
plt.show()
```

3.2 Research Question 2

How can leak detection systems be designed to increase their sensitivity to low-pressure leaks that occur far away from the inlet sensors, and which cannot be detected by these inlet sensors.

3.2.1 Research Question 2: Methodology

The methodology deployed for this research question includes the following

1. Simulate single leaks at 2km, 4km, 10km 16km and 18km on the 20km horizontal pipeline using PIPESIM
2. Determine the impact of the pipeline leaks at those locations on the inlet and outlet pressure sensors
3. Determine the impact of leak location on the detection accuracy of the inlet pressure sensor.
4. Determine the sensor location with the greatest pressure value for leaks farthest away from the inlet of the pipeline

3.2.2 Experimental Processes for Research Question 2

A 20km horizontal pipeline network with a leak location at the midpoint was implemented using PIPESIM.

(a) Generation of single leaks at 5 locations on the pipeline

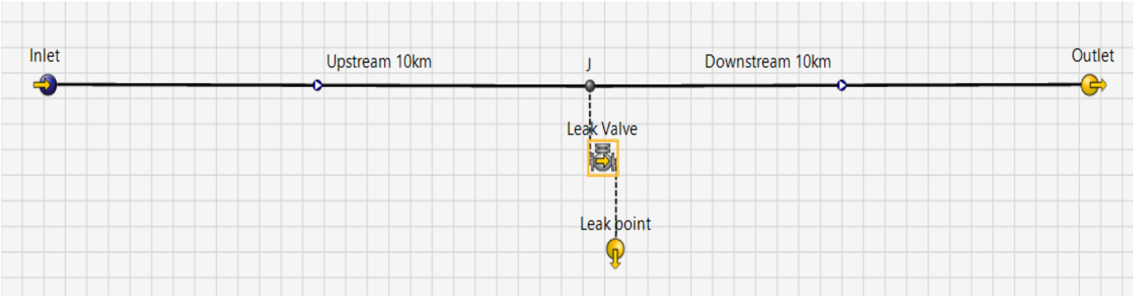


Figure 3.10 Sequential leaks at multiple locations

The pressure profile for the 5 different simulations are plotted and compared with the pressure profile for the no leak case. Figure 3.10 shows the experimental process with sequential leaks at 2km, 4km, 10km 16km and 18km on the 20km horizontal pipeline. The instance shown in the Figure 3.10 is at the 10 km point.

(b) Sensor Sensitivity Analysis

The sensor sensitivity analysis is undertaken to determine the leak location with the greatest pressure value at the inlet pressure sensor

The predominant sensor locations for pipeline leak detection are either at the Inlet and the outlet positions. While this is suitable for ease of monitoring and deployment, the possibility of leak induced pressure variations being lost exists and this is due to the distance between the leak point and the inlet.

(c) Algorithm for Dataset Acquisition and leak detection

Initialization

1. *Set inlet pressure Set_**Pi** (This also represents the Inlet valve status)*
2. *Read Inlet Pressure as **Pi** and record Time **Ti***
3. *Read Outlet Pressure as **Po** and record time **To***
4. *Read Midpoint Sensor Pressure as **Ps** and Record Time **Ts** (Actual Pressure Value)*
5. *Determine Pipeline Pressure Profile model as PPPm*
6. *Calculate pressure at midpoint sensor(**Psm**) using PPPm (expected pressure value)*

Dataset Generation and Leak Detection

7. *Set Pressure Threshold as Thresh*
8. *if (Ps+Thresh) = (Psm +Thresh) and Pi = Set_Pi*
   Store Ps = Dataset

Label Ps = NoLeak
else
9. *if (Ps+Thresh) < (Psm +Thresh) and Pi = Set_Pi*
   store Ps = dataset
   label = leak
10. *if (Ps+Thresh) < (Psm +Thresh) and Pi < Set_Pi*
11. *Determine Pipeline Pressure Profile model as PPPm with new Pi=Set_Pi*
12. *if (Ps+Thresh) = (Psm +Thresh) and Pi = Set_Pi*
   Store Ps = Dataset
   Label Ps = NoLeak
   else
13. *if (Ps+Thresh) < (Psm +Thresh) and Pi = Set_Pi*
   store Ps = dataset
   label = leak

Chapter 4

Results

The results obtained from the simulation experiments are presented in this chapter. The results are presented under the different research questions.

4.0 Results for Pipeline leak and No leak Data generation

4.1 Results for Research Question 1

(a) Pipeline No leak Datasets

The result of the simulation shows the pressure profile of the pipeline. The data generated from the experiments are shown in Table 4.1, while the pressure profile is shown in Figure 4.1

Table 4.1 Pipeline pressure profile data

Pipeline length (km)	Pressure (psia)
0	1000
0.6096	987.5951
0.9999878	979.533
1.609588	966.7574
2.000006	958.4541
2.609606	945.2971
2.999994	936.7461
3.609594	923.1939
4.000012	914.3839
4.609612	900.4187
5	891.3382
5.6096	876.9379

5.999988	867.5703
6.609588	852.7071
7.000006	843.0319
7.609606	827.6715
7.999994	817.6658
8.609594	801.7712
9.000012	791.4188
9.609612	774.973
10	764.2524
10.6096	747.201
10.99999	736.0718
11.60959	718.3456
12.00001	706.7571
12.60961	688.2688
12.99999	676.1598
13.60959	656.7977
14.00001	644.0849
14.60961	623.7034
15	610.2816
15.6096	588.6862
15.99999	574.4088
16.60959	551.331
17.00001	535.992
17.60961	511.045
17.99999	494.344
18.60959	467.1713
19.00001	449.1061
19.60961	419.7055
20	400

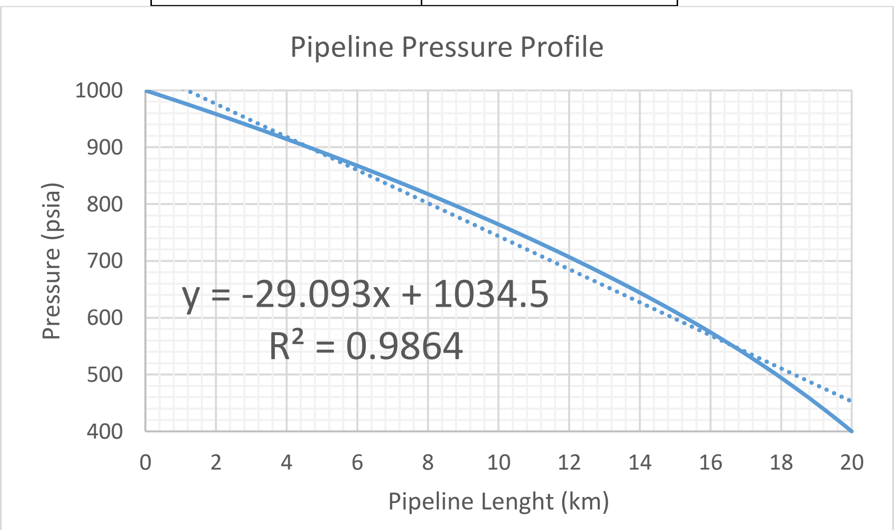


Figure 4.1 Pressure profile of the line with no leaks

The model for the pipeline pressure profile is given in the equation 4.1

$$y = -29.093x + 1034.3 \qquad 4.1$$

From the linear model of the pipeline pressure profile, the pressure values for the different portions of the line can be generated using the pressure profile model. These values can also be read from Table 4.1.

Generating the first 500 variations of the pressure data at the selected sensor location (J) which is at the middle of the pipeline (10km), the RAND function (Random number generator) of python with a peak-to-peak pressure variation of 10 psi around each point on the pipeline is used. The sensor data is shown in Table 4.2.

Table 4.2 Sensor Data

Sensor Data	Parameter
Sensor Location	10km
Pressure at Sensor Location	764psia (rounded to nearest integer)
Inlet pressure	1000psia
Pressure fluctuation value	5psia
Pressure range at sensor location	759 -769psi

The time series data stream at the sensor location (J) is generated and shown below.

[759, 761, 762, 766, 763, 764, 766, 765, 767, 767, 762, 765, 764, 759, 767, 760, 766, 768, 764, 766, 759, 759, 761, 763, 768, 769, 768, 759, 759, 764, 760, 767, 766, 761, 764, 767, 759, 766, 767, 764, 764, 765, 760, 760, 765, 761, 768, 761, 767, 765, 767, 767, 768, 764, 769, 763, 768, 760, 765, 768, 759, 761, 767, 764, 760, 767, 760, 759, 766, 762, 764, 762, 765, 764, 760, 760, 764, 761, 767, 760, 762, 765, 768, 760, 768, 762, 762, 768, 761, 760, 766, 767, 760, 762, 768, 761, 763, 763, 769, 762, 761, 761, 764, 767, 762, 769, 768, 765, 761, 769, 769, 768, 769, 768, 762, 761, 769, 762, 760, 761, 764, 760, 762, 761, 759, 759, 759, 764, 764, 764, 763, 759, 769, 761, 767, 769, 767, 760, 765, 766, 760, 765, 763, 767, 764, 767, 762, 765, 759, 767, 765, 759, 763, 764, 765, 760, 763, 759, 766, 761, 766, 762, 759, 767, 762, 768, 760, 762, 759, 766, 761, 768, 761, 764, 761, 761, 768, 764, 764, 763, 766, 764, 767, 767, 765, 767, 761, 759, 769, 768, 763, 765, 759, 764, 764, 761, 767, 762, 761, 769, 760, 763, 761, 766, 766, 759, 766, 764, 761, 769, 762, 766, 763, 765, 762, 767, 762, 763, 769, 763, 763, 759, 761, 759, 766, 765, 768, 767, 762, 766, 766, 759, 760, 768, 765, 762, 762, 762, 765, 764, 759, 764, 764, 762, 768, 767, 759, 768, 766, 762, 760, 763, 767, 763, 759, 766, 766, 768, 765, 761, 766, 762, 761, 769, 759, 766, 767, 767, 763, 761, 766, 761, 760, 767, 765, 762, 762, 762, 768, 761, 765, 768, 763, 759, 767, 764, 763, 759, 759, 769, 769, 761, 760, 765, 767, 763, 761, 766, 765, 767, 761, 765, 768, 766, 764, 765, 763, 768, 769, 764, 759, 768, 763, 762, 767, 763, 765, 767, 767, 766, 759, 766, 769, 766, 762, 759, 763, 764, 768, 766, 767, 764, 760, 764, 768, 765, 760, 767, 761, 763, 764, 765, 761, 768, 765, 760, 769, 768, 767, 769, 760, 761, 769, 763, 764, 762, 759, 764, 765, 759, 769, 761, 759, 764, 765, 759, 762, 764, 762, 768, 760, 765, 768, 761, 763, 765, 765, 767, 762, 759, 767, 765, 759, 766, 763, 769, 759, 759, 763, 765, 768, 767, 765, 762, 765, 761, 767, 760, 769, 767, 765, 764, 763, 764, 761, 769, 766, 762, 769, 762, 769, 761, 762, 759, 760, 763, 765, 765, 768, 765, 769, 763, 765, 760, 769, 764, 767, 763, 761, 761, 768, 762, 766, 767, 761, 760, 759, 760, 762, 762, 761, 762, 761, 759, 759, 768, 762, 768, 760, 767, 760, 762, 763, 762, 765, 759, 759, 767, 765, 759, 767, 768, 767, 763, 762, 765, 769, 761, 769, 768, 762, 763, 764, 769, 766, 764, 763, 760, 764, 761, 767, 762, 764, 762, 761, 760, 769, 764, 765, 762, 761, 759, 768, 765, 760, 759, 768, 763, 764, 759]

Plotting the datapoints, the following timeseries wave form is generated.

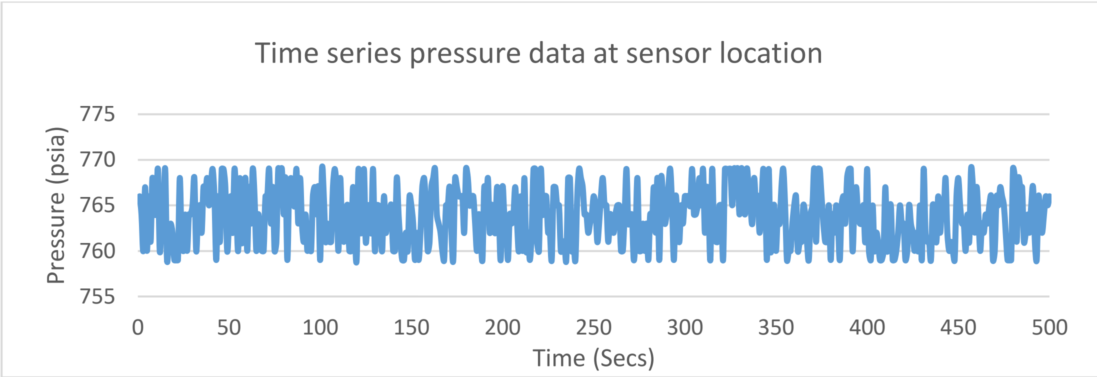


Figure 4.2 Time series Dataset for the No leak condition of the pipeline

Figure 4.2 is the time series dataset for the no leak case of the experiments. This dataset is for a point 10km from the inlet of the pipeline where the pipeline inlet pressure is 1000psi. The model for the pressure profile is used to determine the expected pressure at the sensor location whenever there is a known variation in the pressure at the inlet of the pipeline.

(b) Simulation of a single point leak on the horizontal pipeline using PIPESIM

Pipeline leaks are characterized by sudden drop in the in the pipeline pressure. The ultimate goal of the leak is to equalize the pressure in the pipe with that in the environment as such the leak will continue until this equalization is achieved. The results of the simulation of a 740-psi leak at 10km on the pipeline is shown in Table 4.3. The graph of the pressure profile with the leak is presented superimposed with the no leak case to show the impact of the leak on the pressure profile. This is shown in Figure (s) 4.3 and 4.4.

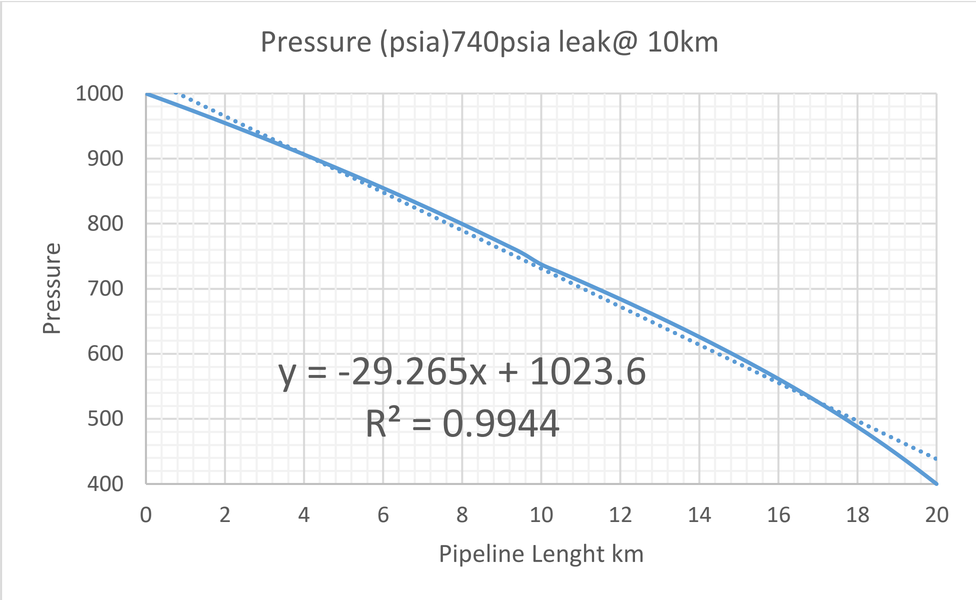


Figure 4.3 Pressure profile of line with One leak at 10km

Figure 4.3 shows the pipeline profile where the line has one leak located at the midpoint. The model of the profile is shown in figure 4.3.

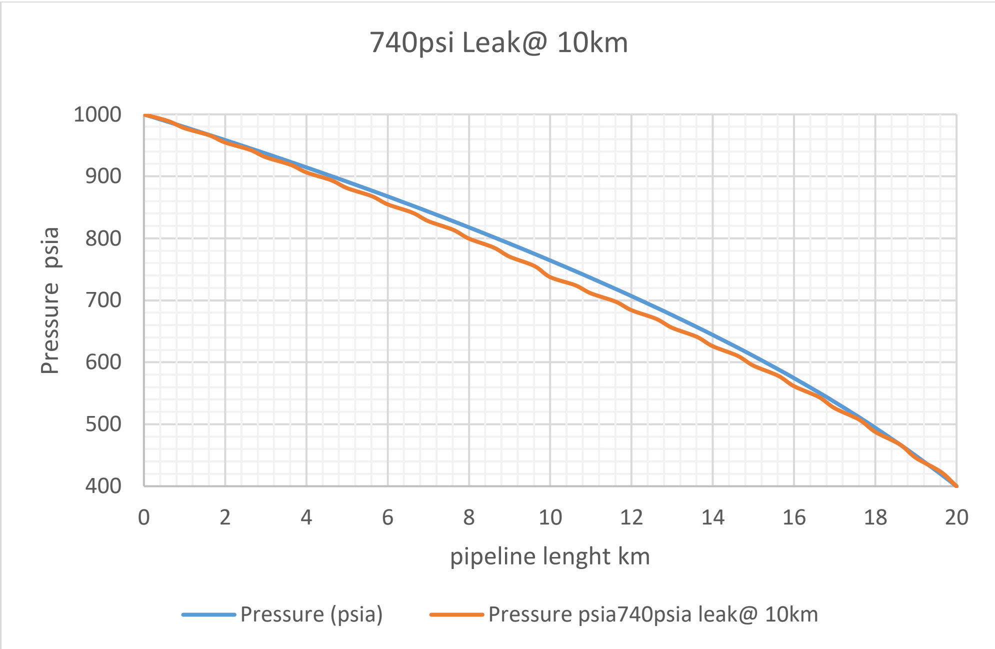


Figure 4.4 Pressure profile of No leak and One leak simulation experiment

Figure 4.4 shows the superposition of the leak profile and the no leak profile.

The Leak and No-Leak pressure profile data is shown in Table 4.3.

Table 4.3 Leak and No leak pressure profile

Pipeline length (km)	Pressure (psia)	Pressure psia740psia leak@ 10km
0	1000	1000
0.6096	987.5951	988.9033
0.9999878	979.533	977.6371
1.609588	966.7574	966.1977
2.000006	958.4541	954.5832
2.609606	945.2971	942.789
2.999994	936.7461	930.8109
3.609594	923.1939	918.6442
4.000012	914.3839	906.2826
4.609612	900.4187	893.7218
5	891.3382	880.9543
5.6096	876.9379	867.9727
5.999988	867.5703	854.7687
6.609588	852.7071	841.3323
7.000006	843.0319	827.6548
7.609606	827.6715	813.724
7.999994	817.6658	799.5311
8.609594	801.7712	785.0765
9.000012	791.4188	770.3462
9.609612	774.973	755.3268
10	764.2524	737.3895
10.6096	747.201	724.4424
10.99999	736.0718	711.2452
11.60959	718.3456	697.7834
12.00001	706.7571	684.0431
12.60961	688.2688	670.0064
12.99999	676.1598	655.6538
13.60959	656.7977	640.963
14.00001	644.0849	625.9081
14.60961	623.7034	610.4619
15	610.2816	594.5902
15.6096	588.6862	578.2544
15.99999	574.4088	561.4093
16.60959	551.331	544.0004
17.00001	535.992	525.9662
17.60961	511.045	507.2286
17.99999	494.344	487.6934
18.60959	467.1713	467.2432
19.00001	449.1061	445.7529
19.60961	419.7055	423.3581
20	400	400

(c) Simulation of a progressive leak at single point leak on the horizontal pipeline

The simulation was done to show the gradual reduction of the leak from 740psi to 350psi, The results shown in Table 4.4 indicate a pipeline failure which is tending to a pipeline burst. From the experiments, there is a fluid reversal from the outlet back to the inlet when the pressure at the leak point drops below the outlet pressure. This pressure profile is shown in Figure 4.5.

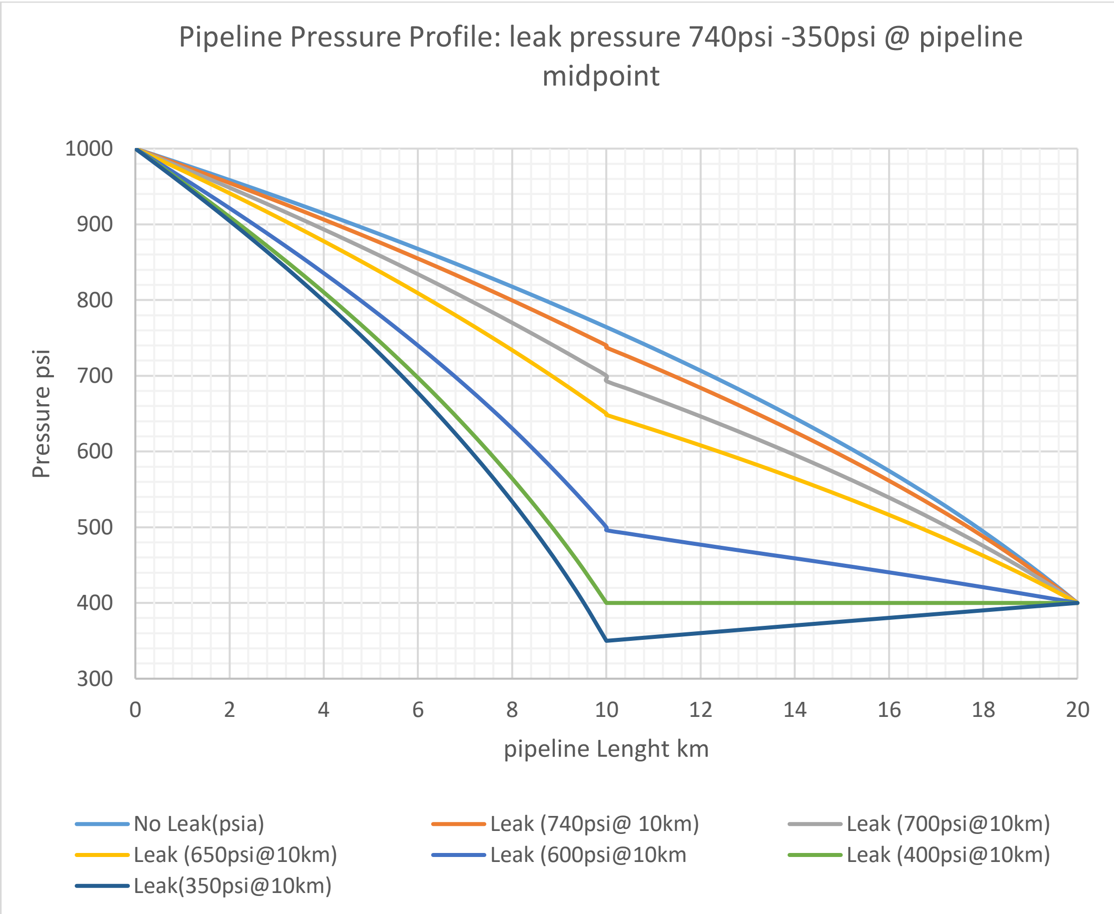


Figure 4.5 Progressive pressure drop at a pipeline leak location (midpoint of the pipeline)

Table 4. 4 Pipeline Leak pressure profile (Leak at leak point= 740 psi to 350psi)

Distance (km)	No Leak(psia)	Leak (740psia @10km)	Leak (700psia@ 10km)	Leak (650psia @10km)	Leak (600psia@ 10km	Leak (400psia@1 0km)	Leak (350psia@1 0km)
0	1000	1000	1000	1000	1000	1000	1000
0.4999939	989.8419	988.9033	987.4052	985.6255	980.9052	978.1507	976.915
0.9999878	979.5329	977.6371	974.6092	971.0091	961.4436	955.8499	953.3376
1.500012	969.0705	966.1977	961.606	956.1415	941.5919	933.0623	929.2261
2.000006	958.4539	954.5832	948.3916	941.0148	921.3275	909.7521	904.5371
2.5	947.6799	942.789	934.9586	925.6176	900.6209	885.8732	879.2156
2.999994	936.7459	930.8109	921.2997	909.9379	879.4395	861.3737	853.1986
3.499988	925.6485	918.6442	907.4069	893.9624	857.746	836.1928	826.4118
4.000012	914.3836	906.2826	893.2704	877.6751	835.496	810.2572	798.7647
4.500006	902.9485	893.7218	878.8817	861.0611	812.642	783.4847	770.1587
5	891.338	880.9543	864.2288	844.1004	789.1241	755.8072	740.529
5.499994	879.5471	867.9727	849.2991	826.7715	764.8988	727.1509	709.75
5.999988	867.5702	854.7687	834.0784	809.05	739.9229	697.3825	677.6469
6.500012	855.4003	841.3323	818.5499	790.9098	714.1096	666.3334	644.0476
7.000006	843.0319	827.6548	802.6978	772.3464	687.3587	633.8072	609.0308
7.5	830.4567	813.724	786.5151	753.332	659.5418	599.8538	572.4151
7.999994	817.6659	799.5311	769.9918	733.8304	630.5004	564.3755	533.9062
8.499988	804.6531	785.0765	753.1058	713.8	600.0347	527.0969	493.1102
9.000012	791.419	770.3462	735.8308	693.1904	568.1931	487.6493	449.4758
9.500006	777.9562	755.3268	718.1402	671.9469	534.933	445.5306	402.1994
10	764.2528	737.3895	693.1238	648.3308	496.1799	400.0009	350
10.49999	750.2961	724.4424	681.7738	638.5673	491.2167	400.0009	352.565
10.99999	736.0723	711.2452	670.2022	628.6129	486.3696	400.0009	355.1229
11.50001	721.5648	697.7834	658.3977	618.4593	481.6223	400.0009	357.6741
12.00001	706.7578	684.0431	646.3497	608.0991	476.9616	400.0009	360.2182
12.5	691.6305	670.0064	634.0443	597.5219	472.3756	400.0009	362.7554
12.99999	676.1606	655.6538	621.4663	586.7169	467.8539	400.0009	365.2858
13.49999	660.3225	640.963	608.5988	575.6717	463.3877	400.0009	367.8094
14.00001	644.086	625.9081	595.4215	564.3719	458.9301	400.0009	370.3264
14.50001	627.4193	610.4619	581.9143	552.8032	454.4075	400.0009	372.8365
15	610.2828	594.5902	568.051	540.947	449.8184	400.0009	375.3399
15.49999	592.6309	578.2544	553.8029	528.7831	445.1619	400.0009	377.8365
15.99999	574.4101	561.4093	539.1365	516.2884	440.4373	400.0009	380.3263
16.50001	555.5558	544.0004	524.0118	503.4353	435.6431	400.0009	382.8096
17.00001	535.9935	525.9662	508.3847	490.1943	430.7784	400.0009	385.286
17.5	515.6287	507.2286	492.1998	476.5285	425.8415	400.0009	387.7557
17.99999	494.3456	487.6934	475.3916	462.3954	420.8304	400.0009	390.2185
18.49999	472.147	467.2432	457.8801	447.7442	415.7433	400.0009	392.6744
19.00001	449.107	445.7529	439.5657	432.513	410.5773	400.0009	395.1234
19.50001	425.1087	423.3581	420.3261	416.6293	405.3306	400.0009	397.5653
20	400	400	400	400	400	400.0009	400

(d) Generation of Pipeline Leak Datasets

The simulated time series datasets for the pipeline leaks and the waveform for the simulated leak are shown in Figure 4.6

[763, 766, 767, 759, 768, 766, 764, 762, 766, 767, 769, 764, 766, 769, 765, 763, 765, 766, 768, 769, 767, 765, 765, 766, 766, 767, 768, 760, 768, 760, 766, 760, 760, 767, 763, 759, 763, 759, 760, 765, 765, 759, 769, 764, 762, 765, 767, 759, 760, 763, 759, 759, 765, 765, 764, 767, 766, 762, 760, 762, 764, 768, 764, 768, 766, 762, 766, 765, 761, 764, 761, 769, 765, 766, 760, 760, 765, 760, 762, 764, 768, 760, 766, 762, 767, 764, 769, 764, 765, 761, 764, 764, 769, 766, 768, 764, 767, 763, 759, 766, 768, 759, 760, 766, 759, 767, 766, 763, 764, 768, 765, 764, 761, 766, 761, 760, 763, 765, 760, 766, 769, 763, 767, 768, 765, 764, 762, 764, 766, 764, 761, 769, 769, 764, 766, 762, 763, 765, 765, 766, 761, 767, 763, 759, 759, 761, 764, 765, 762, 761, 766, 760, 760, 760, 769, 759, 768, 763, 763, 761, 764, 763, 762, 765, 769, 766, 766, 759, 765, 769, 766, 759, 769, 759, 761, 761, 762, 760, 765, 762, 762, 762, 759, 767, 761, 767, 767, 766, 765, 765, 766, 761, 767, 769, 767, 768, 764, 761, 768, 761, 762, 761, 766, 760, 769, 762, 765, 768, 762, 764, 761, 764, 765, 760, 766, 759, 765, 759, 765, 761, 769, 767, 764, 765, 765, 762, 762, 767, 766, 769, 764, 764, 761, 763, 763, 769, 764, 765, 759, 760, 766, 767, 761, 766, 761, 759, 762, 766, 763, 769, 760, 761, 763, 765, 765, 764, 766, 767, 765, 761, 768, 765, 760, 766, 768, 763, 763, 762, 760, 763, 764, 766, 765, 759, 767, 764, 764, 761, 761, 760, 759, 759, 765, 768, 760, 760, 763, 764, 769, 766, 760, 764, 760, 764, 765, 766, 767, 764, 769, 765, 768, 768, 763, 764, 761, 763, 762, 761, 768, 769, 760, 769, 763, 767, 769, 764, 768, 767, 768, 768, 766, 760, 769, 765, 763, 760, 759, 765, 768, 768, 764, 765, 760, 766, 764, 760, 769, 761, 766, 761, 768, 767, 763, 763, 764, 765, 768, 767, 767, 763, 762, 767, 767, 759, 768, 759, 765, 761, 762, 766, 764, 761, 760, 769, 761, 762, 768, 759, 763, 764, 768, 765, 767, 767, 763, 763, 769, 759, 769, 763, 760, 761, 769, 763, 764, 760, 759, 768, 769, 762, 760, 760, 768, 768, 768, 761, 769, 765, 765, 765, 759, 760, 759, 760, 759, 759, 762, 765, 769, 764, 764, 760, 759, 765, 759, 763, 764, 767, 764, 765, 766, 764, 762, 762, 760, 768, 764, 769, 762, 759, 769, 760, 763, 764, 759, 768, 761, 764, 767, 768, 762, 759, 759, 767, 768, 766, 764, 761, 761, 759, 760, 762, 764, 768, 766, 761, 759, 768, 769, 759, 765, 762, 760, 760, 763, 759, 761, 765, 763, 768, 763, 767, 762, 759, 767, 760, 769, 764, 760, 762, 764, 769, 760, 762, 761, 759, 769, 762, 764, 763, 764, 763, 768, 769, 763, 765, 769, 760, 764, 766, 687, 683, 687, 686, 683, 685, 690, 683, 681, 685, 749, 749, 747, 749, 741, 750, 743, 744, 749, 743, 742, 741, 750, 745, 740, 741, 750, 750, 750, 747, 746, 741, 750, 743, 748, 741, 741, 747, 744, 750, 745, 743, 740, 748, 748, 750, 746, 747, 746, 746, 741, 740, 740, 740, 748, 750, 747, 741, 743, 748, 744, 749, 740, 747, 743, 742, 750, 748, 746, 743, 745, 749, 750, 747, 744, 747, 745, 745, 745, 745, 742, 748, 749, 750, 743, 746, 741, 746, 748, 742, 741, 742, 745, 748, 746, 743, 746, 745, 742, 741, 750, 744, 743, 747, 744, 750, 742, 741, 740, 743, 749, 740, 743, 747, 743, 748, 744, 746, 743, 743, 740, 741, 740, 743, 742, 744, 750, 745, 742, 748, 750, 741, 740, 740, 748, 749, 743, 746, 741, 744, 740, 750, 748, 740, 749, 748, 747, 747, 749, 740, 749, 740, 743, 743, 745, 746, 740, 750, 749, 742, 742, 742, 747, 747, 742, 743, 742, 748, 741, 745, 742, 744, 742, 748, 741, 745, 741, 743, 741, 749, 742, 744, 743, 740, 744, 740, 749, 749, 747, 748, 750, 748, 745, 744, 742, 741, 742, 745, 745, 742, 740, 750, 744, 742, 745, 750, 743, 749, 748, 744, 747, 748, 748, 741, 740, 750, 747, 743, 743, 747, 740, 743, 741, 746, 742, 750, 747, 750, 744, 750, 744, 744, 750, 745, 744, 740, 743, 746, 743, 745, 747, 744, 743, 748, 747, 743, 740, 747, 750, 744, 742, 741, 748, 746, 744, 743, 742, 741, 746, 744, 744, 750, 746, 749, 740, 744, 741, 748, 750, 744, 742, 748, 740, 750, 745, 749, 740, 740, 741, 743, 745, 748, 747, 749, 742, 743, 742, 741, 745, 743, 745, 748, 744, 741, 750, 748, 743, 746, 749, 749, 740, 748, 746, 743, 744, 750, 743, 749, 746, 748, 741, 749, 744, 747, 748, 741, 742, 749, 745, 749, 750, 747, 741, 750, 748, 743, 743, 749, 741, 743, 748, 742, 741, 749, 748, 749, 744, 741, 740, 750, 746, 748, 743, 750, 745, 744, 746, 744, 741, 749, 745, 747, 746, 749, 747, 749, 750, 750, 747, 750, 750, 749, 744, 749, 741, 742, 749, 744, 743, 749, 750, 750, 745, 750, 750, 744, 748, 740, 743, 747, 747, 745, 740, 745, 741, 742, 740, 749, 741, 750, 745, 748, 749, 749, 741, 745, 745, 742, 742, 740, 743, 748, 747, 745, 746, 750, 745, 747, 747, 743, 745, 747, 747, 748, 749, 742, 740, 742, 742, 750, 741, 740, 743, 742, 742, 748, 743, 741, 746, 748, 743, 743, 741, 750, 741, 745, 742, 742, 741, 747, 743, 748, 747, 748, 741, 747, 746, 748, 747, 741, 750, 741, 744, 743, 750, 740, 750, 750, 744, 744, 742, 740, 740, 746, 748, 741, 749, 745, 740, 746, 750, 742, 743, 740, 746, 746, 741, 750, 749, 747, 741, 745, 742, 740, 750, 742, 746, 749, 748, 742, 748, 750, 743, 747, 748, 743, 743, 742, 743, 747, 749, 740, 746, 747, 750, 746, 741, 741, 745, 746]

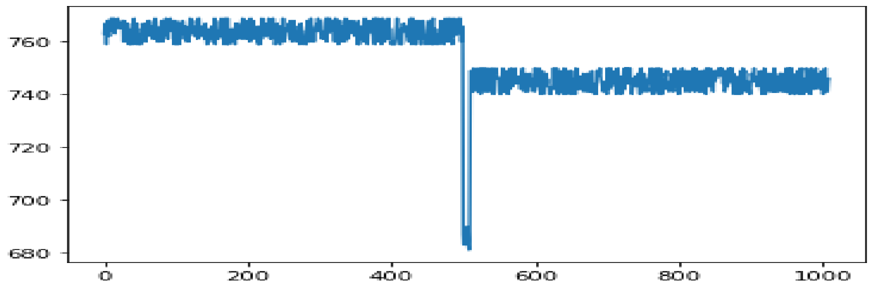


Figure 4.6 Time series waveform for pipeline leak dataset

Figure 4.6 shows the time series waveform of the simulated leak datasets

4.2 Results for Research Question 2

(a) Simulation of multiple leaks on the horizontal pipeline

Table 4.5 shows the data generated from the simulation of pipeline leaks at 2km, 4km, 16km and 18km from the inlet. The pressure reading with a leak at the 10km had been earlier simulated and results shown in Table 4.3.

Table 4.5 Pipeline Pressure leak dataset for single leaks at different points

Distance (km)	No Leak Pressure psia	Leak 850psi@2km	Leak 800psi@4km	Leak 450psi@16km	Leak 450psi@18km
0	1000	1000	1000	1000	1000
0.3333293	993.2441	976.1962	984.7504	991.8652	992.8442
0.6666586	986.4214	951.9658	969.2893	983.6453	985.6165
0.9999878	979.5314	927.2737	953.6059	975.3388	978.3162
1.333348	972.573	902.0778	937.6868	966.9437	970.9418
1.666677	965.5468	876.3377	921.5219	958.4599	963.4939
2.000006	958.4514	850	905.0951	949.8851	955.9708
2.000006	958.4514	840.9446	905.0951	949.8851	955.9708
2.333335	951.2862	835.2698	888.3906	941.2176	948.3716
2.666665	944.0503	829.5306	871.3905	932.4555	940.6951
2.999994	936.7428	823.7262	854.0751	923.5967	932.9402
3.333354	929.3621	817.8557	836.4207	914.6384	925.105
3.666683	921.9085	811.9209	818.4055	905.5796	917.1894
4.000012	914.3802	805.9225	800	896.4173	909.1914
4.000012	914.3802	805.9225	803.3449	896.4173	909.1914
4.300012	907.5398	800.469	797.9484	888.0802	901.9213
4.600011	900.637	794.963	792.4986	879.6549	894.5818
4.900026	893.6704	789.4036	786.9947	871.1386	887.1714
5.20001	886.6399	783.7912	781.437	862.5303	879.6897
5.500025	879.543	778.1241	775.8237	853.8256	872.1337
5.800009	872.38	772.4026	770.1553	845.0236	864.5034

6.100023	865.1479	766.625	764.4301	836.1194	856.7955
6.400008	857.847	760.7916	758.6484	827.1117	849.0097
6.700022	850.4741	754.9005	752.8083	817.9954	841.1425
7.000006	843.0293	748.9519	746.9101	808.7685	833.1935
7.300021	835.5092	742.9438	740.9517	799.4269	825.1588
7.600005	827.9137	736.8763	734.9332	789.9745	817.0378
7.90002	820.2392	730.7471	728.8525	780.406	808.8263
8.200004	812.4852	724.5563	722.7097	770.7193	800.5274
8.500019	804.6509	718.3014	716.5022	760.9079	792.1391
8.800003	796.7385	711.9825	710.2301	750.9689	783.6603
9.100017	788.7442	705.5967	703.8907	740.8951	775.0866
9.400002	780.6674	699.144	697.4838	730.683	766.4167
9.700016	772.5039	692.6213	691.0066	720.3243	757.6457
10	764.2529	686.0284	684.4586	709.8145	748.7719
10.3	755.9104	679.3623	677.8372	699.1448	739.7902
10.6	747.4739	672.6216	671.1408	688.3079	730.6972
10.90001	738.9396	665.8037	664.3668	677.2947	721.4882
11.2	730.3057	658.9075	657.5143	666.0974	712.1599
11.50001	721.5667	651.9293	650.5794	654.7034	702.7055
11.8	712.7205	644.8678	643.5611	643.1038	693.1216
12.10001	703.7611	637.719	636.4553	631.2835	683.4006
12.4	694.6859	630.4815	629.2605	619.2311	673.538
12.70001	685.4882	623.1505	621.9721	606.9285	663.5252
12.99999	676.1645	615.7243	614.5885	594.3609	653.3567
13.30001	666.7074	608.1977	607.1044	581.5063	643.0224
13.59999	657.1127	600.5684	599.5176	568.3451	632.5155
13.90001	647.3717	592.8307	591.8225	554.8497	621.8243
14.19999	637.4792	584.9816	584.016	540.994	610.9403
14.50001	627.4253	577.0146	576.0918	526.7424	599.8494
14.79999	617.2033	568.9261	568.0462	512.1088	588.5408
15.10001	606.8018	560.7087	559.872	497.1453	576.9978
15.39999	596.2124	552.3578	551.5647	481.8285	565.2064
15.7	585.4214	543.8648	543.1156	466.1242	553.146
15.99999	574.4182	535.224	534.5193	450	540.7984
15.99999	574.4182	535.224	534.5193	455.1114	540.7984
16.33332	561.9246	525.4385	524.7839	450.8607	526.7117
16.66665	549.1283	515.447	514.8432	446.553	512.2031
16.99998	536.0043	505.2356	504.6836	442.187	497.2394
17.33334	522.5228	494.7877	494.2887	437.761	481.8913
17.66667	508.6534	484.0878	483.6432	433.2743	466.1572
17.99999	494.3565	473.1151	472.7264	428.7248	450
17.99999	494.3565	473.1151	472.7264	428.7248	449.1758
18.33332	479.644	461.847	461.5161	424.1108	441.4379
18.66665	464.5739	450.2574	449.9865	419.4304	433.5321
18.99998	449.1144	438.3163	438.1079	414.6816	425.4479

19.33334	433.2274	425.9876	425.8448	409.8615	417.1729
19.66667	416.8732	413.2323	413.1587	404.9687	408.6955
20	400	400	400	400	400

Single leaks were simulated at different points on the pipeline with a view to determining the impact of the pressure variation from the leak point on the sensitivity of pressure sensors installed at both the Inlet and the outlet points of the pipeline. Figure 4.7 is the leak located 2km from the inlet of the pipeline.

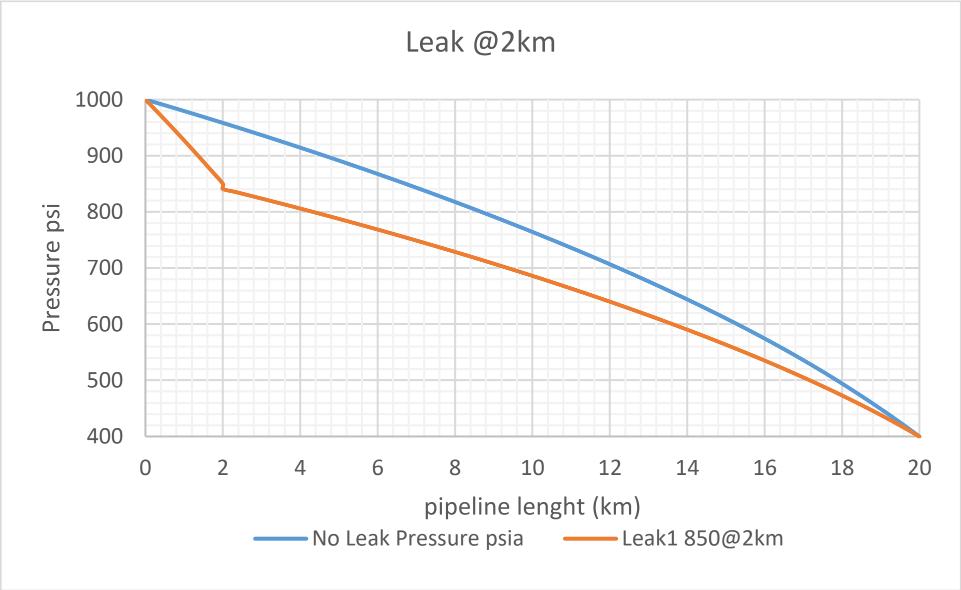


Figure 4.7 Leak 2km from the inlet

Figure 4.8 is the leak 4km from the inlet

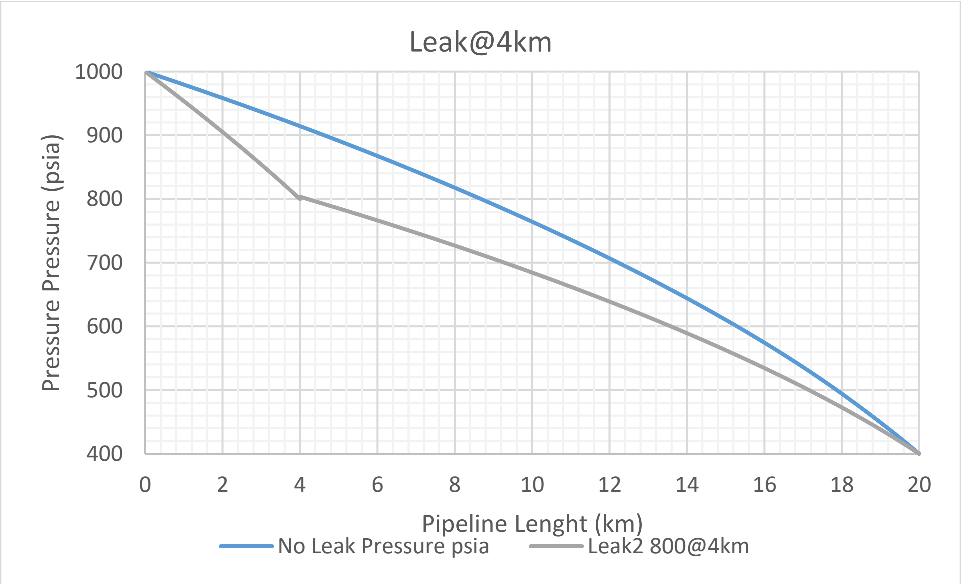


Figure 4.8 Leak 4km from the inlet

Figure 4.9 is the leak at the pipeline midpoint (10km)

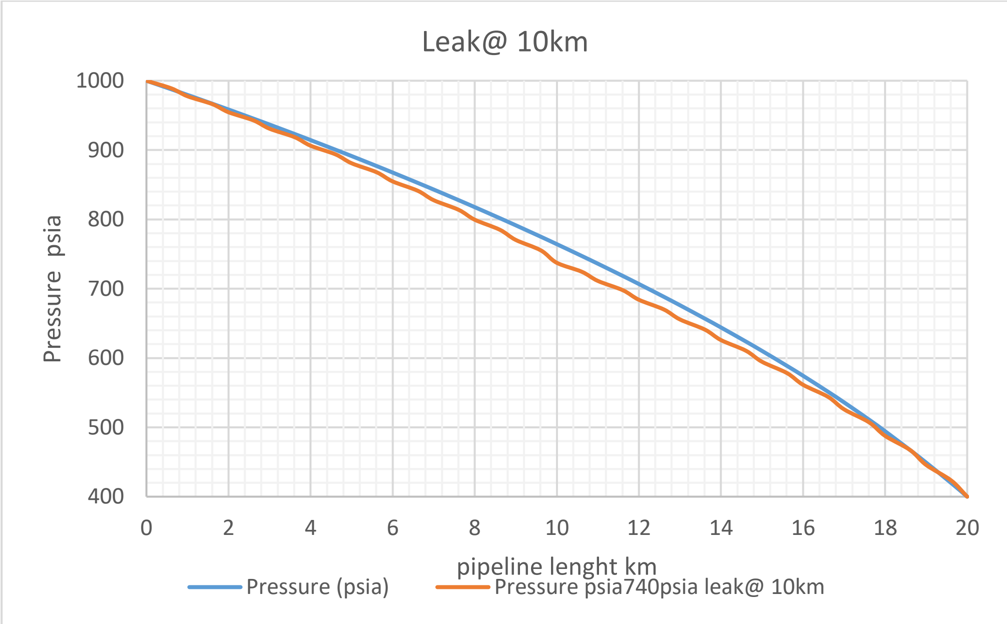


Figure 4.9 Leak at the midpoint (10km)

Figure 4.10 is the leak 4km from the outlet

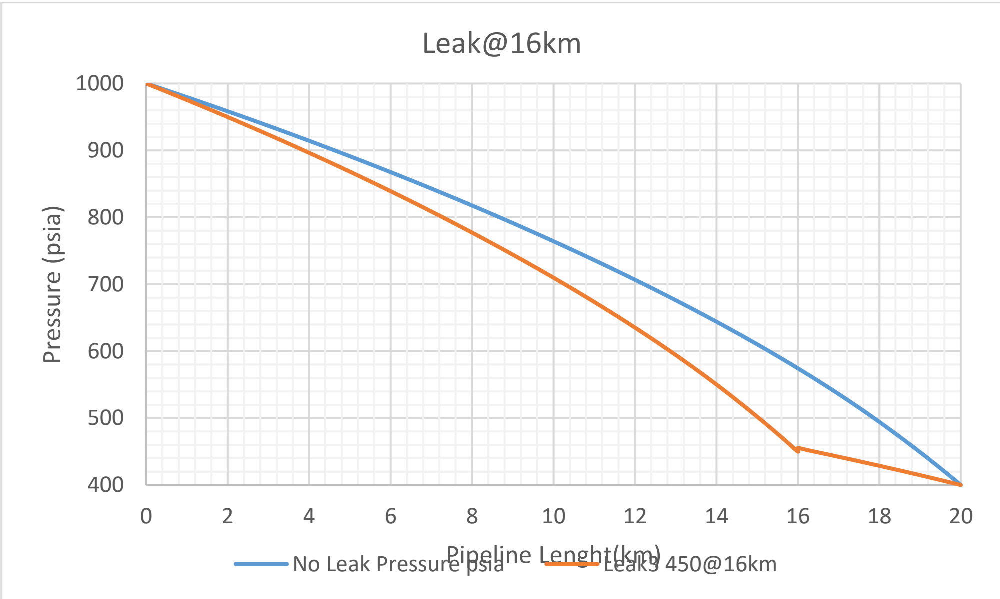


Figure 4.10 Leak 4km from the outlet

Figure 4.11 is the leak located 2km from the outlet

Figure 4.11 Leak 2km from the outlet

Figure 4.12 is the superposition of all the profile of leak points with the No leak pressure profile.

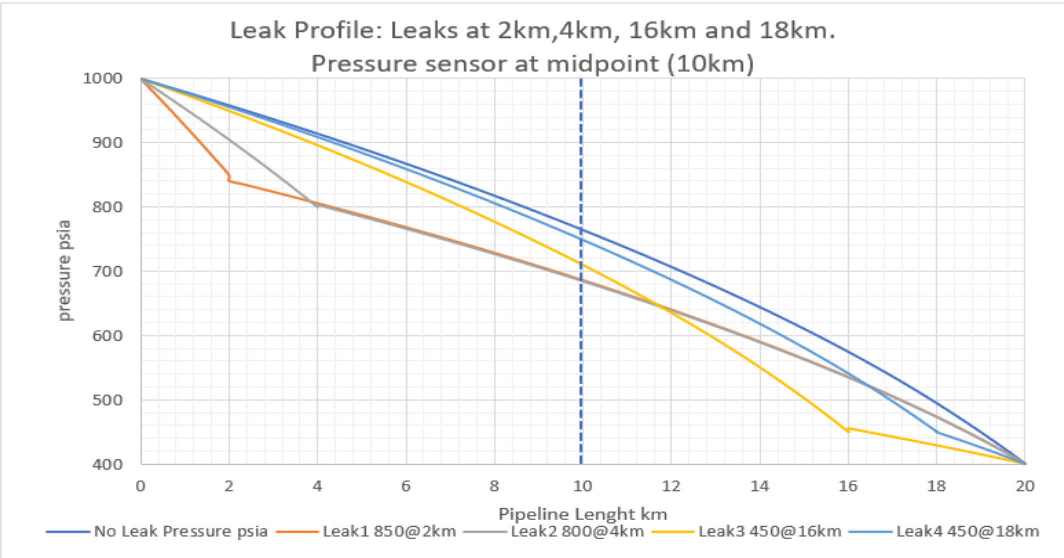


Figure 4.12 Pressure profile for different leak locations

From the results generated from the simulated leaks, the following can be inferred

1. The leaks closest to the Inlet generate the greatest pressure variation both at the inlet and along the pipeline up to the mid-point

2. The leaks closest to the outlet have very negligible pressure variation at the inlet as such the sensors located at the inlet may not be able to sense the pressure variation caused by leaks occurring very close to the outlet.
3. Leaks closest to the outlet are the most difficult to detect by the inlet sensors. Due to the low-pressure value at the outlet, the leak pressure variations would have been absorbed before the variations arrive at the inlet pressure

(b) Sensor Sensitivity Analysis

The predominant sensor locations for pipeline leak detection are either at the Inlet and the outlet positions. While this is suitable for ease of monitoring and deployment, the possibility of omitting pressure variations generated by low volume leaks located far from the inlet pressure sensor exists. The results from the experiments show that leaks occurring close to the inlet have the greatest possibility of detection due to the high pressure at those locations while leaks at the midpoint to the outlet have the highest chance of being undetected due to the low pressure at those locations. Figure 4.13 shows the sensor sensitivity analysis for each of the leak locations and the sensor locations with a sensor installed at the midpoint of the line.

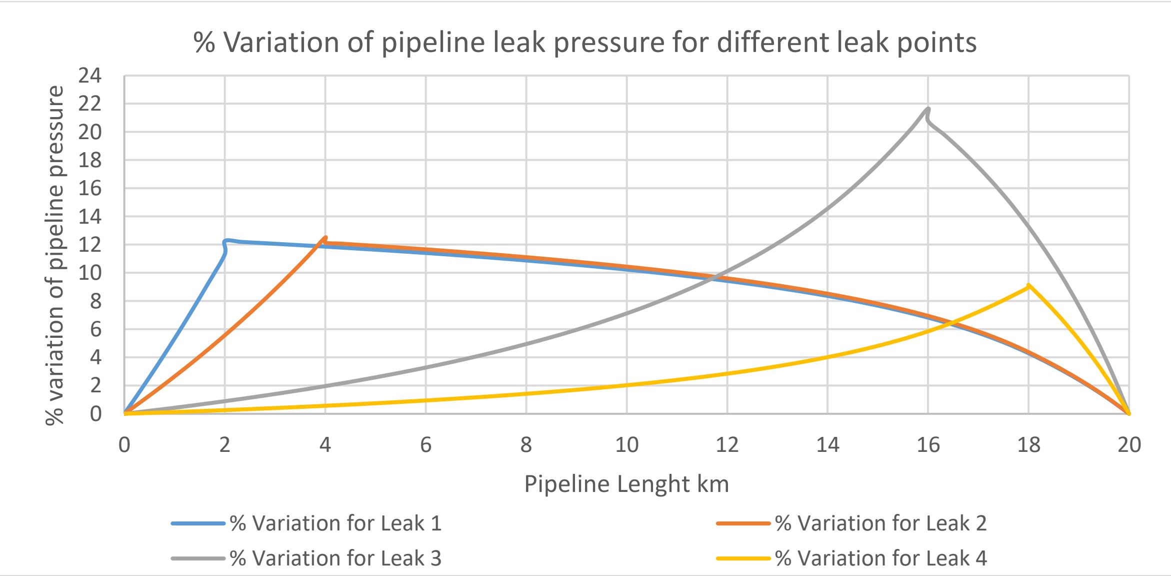


Figure 4.13 Sensor sensitivity analysis

From the results of the pressure variation in Figure 4.13, the pressure variation is highest when the leak is closest to the pressure sensor. Thus, leaks closest to the inlet present a higher pressure variation to the inlet sensor while presenting a lower pressure variation to the opposite sensor. The leak 1 and 2 which are closest to the inlet sensor , presents a higher pressure variation to the inlet sensor and a lower pressure variation to the outlet sensor. The leaks 3 and 4 which are closer to the outlet presents and higher pressure variation to the outlet and a lower pressure variation to the Intel sensor. As sensors age, they lose sensitivity and are not able to pick low pressure leaks which occur further away from the leak. A typical case is that of the leak 3 and 4. The pressure variation at the inlet drops to almost 0 and if the sensors sensitivity has dropped due to age, it will not be able to detect the occurrence of the leak at a 18km on the 20km line. To remedy this situation this research proposes the installation of a pressure sensor at a location midpoint on the pipeline . From the graphs, this sensor located at the midpoint which is 10km has better response to the pressure variations occasioned by leaks that are both close to the inlet and also the outlet. It also has superior performance to leaks located at the midpoint which may be undetectable by the inlet and outlet sensors.

Chapter 5

Discussion

This chapter presents a discussion of the results obtained from the different simulation experiments and the significance of the results in the development of a pipeline leak detection system.

5.0 Discussion

Pipeline leaks are one of the key justifications employed by groups against the continuous exploration and development of the petroleum industry. The significance associated with pipeline leaks is premised on the damage a single leak can cause on the environment if it is not detected and contained quickly. Single spills have been known to discharge millions of gallons of petroleum products into the environment, impacting rivers, farms, and communities. The leaks also impact the finances of the operators in the form of lost production, fines, and clean-up costs. Generally, where the fines range in the 10s of millions of dollars, the cleanup cost extends to the hundreds of millions of dollars while the environmental impact cannot be fully estimated. A review of the challenges associated with leak detection shows that in some cases operators continue to produce from the pipeline even after receipt of the leak alerts. These actions contribute to the increase in the volume of oil spilled into the environment and the actual volume of oil spilled into the environment is only determined after investigations have been undertaken by the regulatory agencies. There is there for a need for a platform for monitoring the Leak Detection Alerts from the operator network which is accessible to the regulator. This will reduce the delays in operator response to pipeline leaks by the operators.

Leak detection is an ongoing research area as the need for a faster LDS cannot be overemphasized. The progression from the manual pipeline right-of-way monitoring to the hardware-dominated exterior systems-based approach and the software or interior-based systems for the development of LDS has resulted in changes in the LDS costs and detection efficiency. While the Hardware-based approach provided better coverage of the entire pipeline network, It is prohibitively expensive to deploy and maintain as such very few operators are willing to deploy them. The software-based approach on the other hand relies on data generated from the LDS and the application of computational pipeline models. While the software-based approach is cheaper in terms of installation cost, the complexity associated with the development is high. Machine learning is a current approach being explored for the development of Leak detection systems as it relies on the pipeline datasets for the development of leak detection algorithms. The key challenge with this approach is the lack of datasets for the development of these algorithms.

A limitation in the existing LDS systems is their inability to detect leaks below their minimum detectable levels. The main factors responsible for keeping the leaks below the minimum levels of the LDS include the leak size and the leak location. Small leaks which can be caused by corrosion failure can go undetected because of the leak size and the flow rate of the leak. The LDS with the lowest minimum leak detection rate is based on the extended real-time transient model which has a minimum leak rate of 0.5% of the nominal flow rate of the line as such leaks which are below this value can continue undetected.

A key challenge addressed by this research is the detection of low pressure/low-volume pipeline leaks that are undetectable by the existing LDS. A review of this challenge shows that the leak-induced pressure is dissipated before getting to the inlet sensor, and the volume of the leak is below the Minimum of the LDS. The most suitable solution for detecting this is the Optic Fiber System but it is very costly to deploy. The findings from the research experiments show that low-volume leaks are below the minimum detection levels of the installed LDS and the perturbations caused by these leaks are within the pressure variation range in the pipeline, the leak-induced pressure variation from the leak point is dissipated before it gets back to the inlet pressure sensor.

Machine learning provides an opportunity for the development of faster and more robust LDS as research has shown that Artificial Neural Networks used in the development of LDS generates a very high detection accuracy for leak detection However, it requires a large volume of datasets which are not readily accessible to the research community. Where the data exist, the Machine learning algorithm is faced with class imbalance challenges as the data will contain more of the No leak cases compared with the leak cases. The discussions on the specific contributions of the research are presented in the following sections.

5.1 Pipeline Leak and No leak Time series Dataset Generation

The use of PIPESIM and python for the generation of time series pressure data facilitates the generation of both the leak and no leak datasets. One key challenge associated with pipeline leak detection research is the unwillingness of operators to share their datasets with the research community. This work has developed a system for generating pipeline datasets through the use of the PIPESIM software. The GIS capability of PIPESIM enables the development of the actual operator pipeline network using the topology of the pipeline right of way. With this software, all that is required is the inlet pressure, the pipeline network topology for the determination of the pipeline pressure profile. The use of the RAND function in python enables the generation of time series pressure at any point on the pipeline. This functionality will also enable the installation of virtual pressure sensors at any point on the line without having to install any physical sensor. This system will enable the generation of pipeline dataset from any point on the line.

The insertion of a choke and a sink also enables the generation of pipeline leak datasets in the network using the PIPESIM simulation and the RAND function. This functionality also enables the simulation of pipeline leaks at any section of the pipeline and observing the impact of the leak on the network. This combination of PIPESIM and the RAND function is suitable for generating pipeline leak and no leak datasets from any portion of the pipeline.

5.2 One class classifier and leak dataset generation

The one-class classification algorithm requires that the dataset be solely of one class and the Machine learning algorithm requires a large volume of datasets which is also a challenge that this research has been able to address. With this framework, researchers can generate datasets from any part of the pipeline and can also generate the required volume of datasets needed for their research. This is an improvement in the current state of the art as it eliminates the need for physical sensors on pipeline networks for the generation of pipeline datasets. This approach is cost-effective and provides a platform for the development of robust leak detection systems using artificial intelligence and machine learning algorithms.

An anomaly in the pipeline pressure data is meant to indicate the presence of a leak. Other parameters not associated with leaks can cause variations in the pipeline pressure thus generating false alarms. It is there required that these non-leak causing activities be identified and isolated from the leak decision making process. In this research, the inlet pressure is monitored and compared with the midpoint pressure value. Variations in the inlet pressure are tracked and correlated with the midpoint pressure sensor and a model of the flow is developed. With this, the midpoint pressure sensor value tracks the inlet pressure values and this reduces the generation of spurious signals or erroneous signals. The algorithm presented in this work enables the leak detection model to track the inlet pressure and ensures that variations in the inlet pressure which are reflected in the midpoint pressure sensor are not classified as leaks. This minimizes the generation of false alarms and also enables the development of a lightweight model for leak detection. This model utilizes only the Inlet pressure and the midpoint pressure sensor for the development of the real time leak detection algorithm.

5.3 Midpoint Sensor Installation on the Pipeline

This research has been able to establish that the possibility of detection for low-volume leaks decreases as the leak point moves away from the inlet sensors. Most leak detection systems rely on the pressure sensors installed at the inlet of the pipeline as such the pressure variation generated by the leak dissipates before it gets to the inlet pressure.

The results obtained from this research show that the installation of a pressure sensor at the midpoint of the pipeline increases the chances of the detection of low-volume leaks which occur at locations where they would be undetectable by the inlet pressure sensor. The midpoint pressure sensor installation also increases the robustness of the LDS as it can counter the loss of sensitivity faced by the inlet pressure sensor and increase the sensitivity of the LDS by enabling the detection of the leak which is lower than the minimum detection levels of the installed LDS.

The result in figures 4.7 to 4.11 shows that the sensors located at the inlet and outlet are most responsive to leaks closest to their locations. For very long lines and small leaks, the pressure variation gets dissipated before it gets to the sensor locations. The introduction of a sensor at the midpoint of the pipeline provides an opportunity for the detection of small pressure variations caused by leaks which are around the midpoint of the pipeline. The introduction of the exception based algorithms in the midpoint sensors enables the batteries serving such sensors to last much longer than is expected as transmission only takes place when the pressure falls outside the expected envelop.

5.4 Leak Detection as a Service (LDaaS). System Overview

One of the finding of this research has been the fact that there is no platform where regulators in the oil and gas industry responsible for oil spill monitoring can access the leak alerts generated from operator pipeline networks. These regulators rely on reports from the operators or reports from the public whenever there is a leak. This gap is the reason why operators can continue production for several hours after a leak has been detected and also underreport or underestimate the volume of oil spilled as a result of their leak. Studies and data shows that the volume of oil spilled into the environment is directly proportional to the time it takes to detect and contain the leak. The leak detection architecture proposed in this research provides an opportunity for the development of a platform where regulators can register and receive the leak alerts from the operator pipeline networks as soon as they are generated. This feature is made possible by the development of the Leak Detection as a Service platform. The Leak Detection as a Service (LDaaS) is a platform that utilizes the pressure from the inlet and the midpoint of the pipeline in a specially developed algorithm to detect pipeline leaks. The data from these two sensors are transmitted to a cloud location and the leak detection algorithm utilizes these datasets to provide real time leak detection services to these operators without the operators having to deploy their own leak detection systems.

The algorithm is designed to be lightweight as it only utilizes pressure data from two sensors as against existing leak detection models which require multiple data sources to determine the occurrence of a leak. This platform will provide a low cost approach to operators especially the marginal operators with low footprints to deploy leak detection systems on their pipelines. It can accommodate multiple operators and ultimately reduce the cost of deploying leak detection systems by eliminating the need for proprietary leak detection systems. The relevant regulatory agencies responsible for oil spill monitoring can also register on the platform to monitor the leak alerts generated from the operator pipeline and evaluate their response to the alerts. The sensors transmitting to the LDaaS platform are designed to be GPS enabled such that their locations can be tracked while the exception based transmission feature integrated into the sensors will enable them to operate for extended durations. The block diagram of the proposed LDaaS system is shown in Figure 4.1. The sensors S1 represents the inlet pressure sensor, the S2 sensor represents the midpoint sensor while the S3 represents the outlet sensor. P1, P2 and P3 represents the inlet pressure, the midpoint pressure and the outlet pressure.

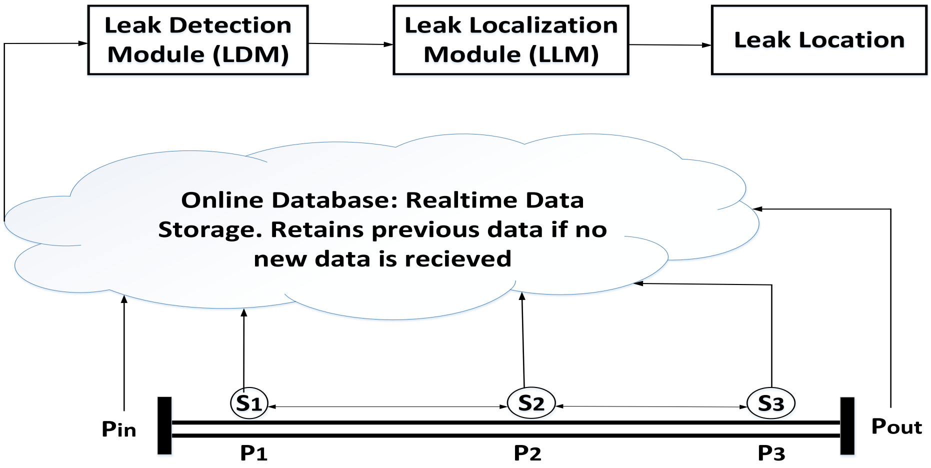


Figure 5.1 Leak Detection as a Service Platform

Multiple sensors from different operators can be managed by one platform with each pipe section identified by the inlet, midpoint and outlet pressure sensors.

Chapter 6

Conclusions

This section presents the contributions of the research, the key conclusions, and recommendations for further research.

6.0 Research contributions

1. The research has been able to develop a framework for generating both the leak and no-leak time series datasets from any location of the pipeline. The use of PIPESIM with its GIS functionality and the RAND function of Python will enable researchers to build a pipeline network using the actual pipeline network topography and simulate leaks at any point of the pipeline.
2. This research has also been able to develop a pipeline leak detection architecture with a pressure sensor at the middle of the pipeline capable of detecting low volume leaks released from small diameter leak orifices. The size of the leak orifice and the pressure of the line determines the volume of the oil released from that orifice. At locations far from the inlet, the pressure is reduced and with a leak from a small leak orifice, the inlet sensor will not be able to detect the leak. The installation of a sensor at the midpoint of the line will enable the leak detection system to detect small leak-induced pressure variations which ordinarily would have been missed by the sensors at the inlet and outlet of the pipeline.

`6.1 Recommendations for further research

The recommendations for further research in this area include the following

1. Development of the leak detections system for subsea pipelines to investigate the impact of underwater transmission of the pressure signals to a cloud location
2. The implementation of this research using flowrate sensors to evaluate and compare the flowrate-based system against the pressure-based system used in this research.
3. The deployment of the leak detection algorithm developed from this research using the One Class Dataset model which tracks the inlet pressure and is able to detect leak-induced pressure variation at the midpoint sensor.
4. The deployment of the Leak Detection as a Service platform comprising three pressure sensors and the lightweight leak detection algorithm developed for real-time leak detection from pipeline networks.

References

Adegboye, M.A., Fung, W.K., Karnik, A., (2019). Recent advances in pipeline monitoring and oil leakage detection technologies: principles and approaches. Journal of Sensors 19 (11), 2548. https://doi.org/10.3390/s19112548.

Akerkar R. (2019) Machine Learning. In: Artificial Intelligence for Business. Springer Briefs in Business. Springer, Cham. https://doi.org/10.1007/978-3-319-97436-1_2

Akinsete, O., Oshingbesan, A., (2019). Leak detection in natural gas pipelines using intelligent models. In: SPE Nigeria Annual International Conference and Exhibition, 5-7 August, Lagos, Nigeria. https://doi.org/10.2118/198738-MS.

Alpandi. A H, Mazeli. A H, Sidek. A, Husin.H, Junin R and Jaafar M Z.(2021) Flow pattern, pressure drop and inclination analysis on liquid-liquid two phase flow of waxy crude oil in pipelines using PIPESIM. IOP Conf. Series: Materials Science and Engineering 1142 (2021) 012008

Araújo, M. V.; Farias Neto, S. R.; Lima, A. G. B.; Luna, F. D. T. Hydrodynamic study of oil leakage in pipeline via CFD. (2014) Advances in Mechanical Engineering, p. 170-178, 2014. DOI:10.1155/2014/170178.

Basu, S., & Meckesheimer, M. (2006). Automatic outlier detection for time series: an application to sensor data. Knowledge and Information Systems, 11, 137-154.

Baumgarten A. North Dakota pipe leaks 1.4 million gallons of saltwater, 170 times more than initially reported. https://www.inforum.com/news/north-dakota/north-dakota-pipe-leaks-1-4-million-gallons-of-saltwater-170-times-more-than-initially-reported Accessed 22 August 2022

Begovich. O,Pizano-Moreno. A , and Besançon. G. (2012). Online implementation of a leak isolation algorithm in a plastic pipeline prototype. Lat. Am. appl. res. vol.42 no.2 Bahía Blanca abr.
http://bibliotecadigital.uns.edu.ar/scielo.php?script=sci_arttext&pid=S0

Bickis Ian. Limits of pipeline leak detection systems hit spotlight again with the Saskatchewan spill (2017). The Canadian Press. https://www.cbc.ca/news/canada/calgary/pipeline-leak-detection-problems-calgary-sensors-ocean-man-first-nation-tundra-energy-1.3958153

Building A Machine Learning Anomaly Detection System Design Principles. Anodot 2017. www.anodot.com

Bergmann. P., Fauser. M., Sattlegger. D., and Steger. C., (2019). Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In CVPR, 2019

Boaz, L., Kaijage, S., Sinde, R., 2014. An overview of pipeline leak detection and location systems. In: Proceedings of the 2nd Pan African International Conference on Science, Computing and Telecommunications (PACT 2014), 14–18 July, Arusha, Tanzania. https://doi.org/10.1109/SCAT.2014.7055147.

Bohorquez, J., A. R. Simpson, and M. F. Lambert. (2018). "Characterization of transient pressure traces due to the effects of different anomalies and features in water pipelines." In Proc., 13th Pressure Surges Conf., 151–169. Cranfield, UK: BHR Group

Bohorquez, J., B. Alexander, A. R. Simpson, and M. F. Lambert. (2020). "Leak detection and topology identification in pipelines using both fluid transients and artificial neural networks." J. Water Resour. Plann. Manage. 146 (6): 04020040. https://doi.org/10.1061/(ASCE) WR.1943-5452.0001187.

Bolotina, I., Borikov, V., Ivanova, V., Mertins, K., Uchaikin, S., (2018). Application of phased antenna arrays for pipeline leak detection. J. Petrol. Sci. Eng. 161, 497–505. https://doi.org/10.1016/j.petrol.2017.10.059

Brody Kutt, , "Using High-Order Prior Belief Predictions in Hierarchical Temporal Memory for Streaming Anomaly Detection" (2018). Thesis. Rochester Institute of Technology

Billaudelle S. and Ahmad, S.(2015) "Porting HTM models to the Heidelberg neuromorphic computing platform," 2015, http:// arxiv.org/abs/150502142

Candelieria, A, Contib. D, Archettia. F. (2014) A graph-based analysis of leak localization in urban water networks. 12th International Conference on Computing and Control for the Water Industry, CCWI2013

Caputo, A. C., and Pelagagge, P. M., (2003). Using Neural Networks to monitoring piping systems. Process Safety Progress 22(2), 119-127.

Carrera. D, Boracchi. G., Foi. A., and Wohlberg. B.,(2015) Detecting anomalous structures by convolutional sparse models. In IJCNN.

Chalapathy, R and Chawla, S. "Deep learning for anomaly detection: a survey," (2019), https://arxiv.org/abs/1901.03407

Chen, Q., Shen, G., Jiang, J., Diao, X., Wang, Z., Ni, L., Dou, Z., (2018). Effect of rubber washers on leak location for assembled pressurized liquid pipeline based on negative pressure wave method. Journal of Process Safety and Environmental Protection 119, 181–190. https://doi.org/10.1016/j.psep.2018.07.023

Chuanbo Cui, Shuguang Jiang, Xinjian He, Kai Wang, Hao Shao, Zhengyan Wu. (2018). Experimental study on the location of gas drainage pipeline leak using cellular automata. Journal of Loss Prevention in the Process Industries.vol 56 pp 68-77

Covas, D., H. Ramos, and A.B. de Almeida, (2005) "Standing wave difference method for leak detection in pipeline systems," *Journal of Hydraulic Engineering,* **131,** 1106-1116.

Colebrook, C.F. (1939) Turbulent Flow in Pipes, with Particular Reference to the Transition Region between the Smooth and Rough Pipe Laws. Journal of the Institution of Civil Engineers, 11, 133-156.

http://dx.doi.org/10.1680/ijoti.1939.13150

Costanzo, F., Morosini, A. F., Veltri, P., and Savi´c, D. (2014). "Model calibration as a tool for leakage identification in WDS: A real case study." Procedia Eng., 89, 672–678.

De Silva, D., Mashford, J. and Burn. S. (2011) Computer aided leak location and sizing in pipe networks. Technical Report 17, Urban Water Security Research Alliance, April 2011.

DeBofsky, Abigail & Xie, Yuwei & Jardine, Tim & Hill, Janet & Jones, Paul & Giesy, John. (2020). Effects of the husky oil spill on gut microbiota of native fishes in the North Saskatchewan River, Canada. Aquatic Toxicology. 229. 105658. 10.1016/j.aquatox.2020.105658.

Dey, P. K. (2004). Oil pipelines. In C. J. Cleveland, *Encyclopedia of energy*. Elsevier Science & Technology.

Department of Justice., (2021). Pipeline Company Sentenced for Largest-Ever Inland Oil Spill. Pipeline Rupture Discovered After 143 Days and Discharge of 29 million Gallons. https://www.justice.gov/opa/pr/pipeline-company-sentenced-largest-ever-inland-oil-spill. Accessed 25th November 2022

Edrisi, A.; Kam S. I.(2013) Mechanistic Leak-Detection Modeling for Single Gas-Phase Pipelines: Lessons Learned from Fit to Field-Scale Experimental Data. Advances in Petroleum Exploration and Development, v.5, n.1, p.22–36, 2013.DOI: 10.3968/j.aped.1925543820130501.1027.

Fiedler, J., (2016). An overview of pipeline leak detection technologies. KHRONE Inc https://asgmt.com/wp-content/uploads/2016/02/004.pdf .

Francis Idachaba, Minou Rabiei,(2021) Current technologies and the applications of data analytics for crude oil leak detection in surface pipelines, Journal of Pipeline Science and Engineering, Volume 1, Issue 4, 2021, Pages 436-451, ISSN 2667-1433, https://doi.org/10.1016/j.jpse.2021.10.001.

Felicity Barringer (2006) Large Oil Spill in Alaska Went Undetected for Days. https://www.nytimes.com/2006/03/15/us/large-oil-spill-in-alaska-went-undetected-for-days.html. Accessed 25th November 2022

Fukushima, K., R. Maeshima, A. Kinoshita, H. Shiraishi, and I. Koshijima. (2000). “Gas pipeline leak detection system using the online simulation method.” Comp. Chem. Eng. 24 (2–7): 453–456. https://doi.org/10 .1016/S0098-1354(00)00442-7.

Fitzpatrick. A Faith., Michel C. Boufadel, Rex Johnson, Kenneth Lee, Thomas P. Graan, Adriana C. Bejarano, Zhenduo Zhu, David Waterman,Daniel M. Capone, Earl Hayter, Stephen K. Hamilton, Timothy Dekker, Marcelo H. Garcia, and Jacob S. Hassan.(2015) Oil-Particle Interactions and Submergence from Crude Oil Spills in Marine and Freshwater Environments—Review of the Science and Future Science Needs. U.S. Geological Survey, Reston, Virginia: 2015

Gao, Y., Liu, Y., Ma, Y., Cheng, X., Yang, J., (2018). Application of the differentiation process into the correlation-based leak detection in urban pipeline networks. Journal of Mechanical Systems and Signal Processing 112, 251–264. https://doi.org/ 10.1016/j.ymssp.2018.04.036.

Görnitz, N, Kloft,M, Rieck,K and Brefeld. U, (2013) “Toward supervised anomaly detection,” Journal of Artificial Intelligence Research, vol. 46, pp. 235–262, 2013.

Gong, J., A. R. Simpson, M. F. Lambert, A. C. Zecchin, Y. I. Kim, and A. Tijsseling. (2013). “Detection of distributed deterioration in single pipes using transient reflections.” J. Pipeline Syst. Eng. Pract. 4 (1):32–40. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000111

Gong, J., M. F. Lambert, S. T. N. Nguyen, A. C. Zecchin, and A. R.Simpson. (2018). “Detecting thinner-walled pipe sections using a spark transient pressure wave generator.” J. Hydraul. Eng. 144 (2): 06017027. https://doi.org/10.1061/(ASCE)HY.1943-7900.0001409

Gong, J., M. F. Lambert, A. R. Simpson, and A. C. Zecchin. (2014). “Detection of localized deterioration distributed along single pipelines by reconstructive MOC analysis.” J. Hydraul. Eng. 140 (2): 190–198. https://doi.org/10.1061/(ASCE)HY.1943-7900.0000806.

Goldstein M, Uchida S (2016) A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. PLoS ONE 11(4): e0152173. doi:10.1371/journal. pone.0152173

Guo,W., L. Soibelman, and J. H. Garrett. (2009). "Visual pattern recognition supporting defect reporting and condition assessment of wastewater collection systems." J. Comput. Civ. Eng. 23 (3): 160–169. https://doi.org /10.1061/(ASCE)0887-3801(2009)23:3(160).

G. Romero-Tapia, M.J. Fuente, Vicenç Puig (2018) Leak Localization in Water Distribution Networks using Fisher Discriminant Analysis, IFAC-PapersOnLine, Volume 51, Issue 24,

Hajian-Tilaki K. (2013)Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian J Intern Med. 2013 Spring;4(2):627-35. PMID: 24009950; PMCID: PMC3755824.

Huang, P., N. Zhu, D. Hou, J. Chen, Y. Xiao, J. Yu, G. Zhang, and H. Zhang. (2018). "Real-time burst detection in district metering areas in water distribution system based on patterns of water demand with supervised learning." Water 10 (12): 1765. https://doi.org/10.3390 /w10121765.

Hamid Masood Khan , Fazal Masud Khan, Aurangzeb Khan, Muhammad Zubair Asghar and Daniyal M. Alghazzawi. (2021) Anomalous Behavior Detection Framework Using HTM-Based Semantic Folding Technique Hindawi Computational and Mathematical Methods in Medicine Volume 2021,

He, G., Liang, Y., Li, Y., Wu, M., Sun, L., Xie, C., Li, F., (2017). A method for simulating the entire leaking process and calculating the liquid leakage volume of a damaged pressurized pipeline. J. Hazard Mater. 332, 19–32. https://doi.org/10.1016/j. jhazmat.2017.02.039.

Henrie, M., P. Carpenter, and R. Nicholas. 2016. Pipeline leak detection handbook. 1st ed. Cambridge, MA: Elsevier.

Hoarau, Q., Ginolhac, G., Atto, A.M., Nicolas, J., (2017). Robust adaptive detection of buried pipes using GPR. J. Signal Process. 132, 293–305. https://doi.org/10.1016/j. sigpro.2016.07.001

Ian, G., Yoshua, B., Aaron, C., (2016). Deep Learning. MIT Press.

Idachaba F.E, Okuns G., Wokoma. E, Awobamise B (2014). *Remote Pipeline Pressure Monitoring Using Low Power Wireless Transmission.* SPE National International Conference and Exhibition **Nigeria NAICE 2014**

Iqmal Irsyad Mohammad Fuad, Mohd Fuad Ngah Demon, and Husiyandi Husni, (2020). Automated Real Time Anomaly Detection Model for Operation and Production Data at Scale ADIPEC 2020.

Juliano, T., J. Meegoda, and D. Watts. (2013). "Acoustic emission leak detection on a metal pipeline buried in sandy soil." J. Pipeline Syst. Eng.Pract. 4 (3): 149–155. https://doi.org/10.1061/(ASCE)PS.1949-1204.0000134.

J. Hawkins, S. Ahmad, and Y. Cui, (2017). Why Does the Neocortex Have Layers and Columns, a Theory of Learning the 3d Structure of the World, bioRxiv, Redwood City, California, USA, 2017.

Jesse Coleman. (2014). Tesoro's Clean-up of Massive North Dakota Oil Pipeline Spill. https://www.greenpeace.org/usa/tesoros-north-dakota-oil-pipeline-spill-cleanup-photos/ Accessed 25th November 2022

Kang Jiheon, Youn-Jong Park, Jaeho Lee, Soo-Hyun Wang, and Doo-Seop Eom (2018). Novel Leakage Detection by Ensemble CNN-SVM and Graph-Based Localization in Water Distribution Systems. *IEEE Transactions on Industrial Electronics, Vol. 65, No. 5*

Kishan G. Mehrotra , Chilukuri K. Mohan, HuaMing Huang. (2017)Anomaly Detection Principles and Algorithms. Springer 2017

Krawczyk, Bartosz. (2016). Learning from imbalanced data: Open challenges and future directions. Progress in Artificial Intelligence. 5. 10.1007/s13748-016-0094-0.

Koza, J.R., Bennett, F.H., Andre, D., Keane, M.A., (1996). Automated design of both the topology and sizing of analog electrical circuits using genetic programming. In: Artificial Intelligence in Design'96. Springer, Dordrecht, pp. 151–170. https://doi.org/10.1007/978-94-009-0279-4_9.

Krithikadatta J. (2014) Normal distribution. J Conserv Dent. 2014 Jan;17(1):96-7. doi: 10.4103/0972-0707.124171. PMID: 24554873; PMCID: PMC3915399.

Lee, P., M. F. Lambert, A. R. Simpson, J. P. Vitkovsky, and J. A. Liggett. (2006). "Experimental verification of the frequency response method for pipeline leak detection." J. Hydraul. Res. 44 (5): 693–707. https://doi .org/10.1080/00221686.2006.9521718

Lee, P., J. P. Vítkovský, M. F. Lambert, A. R. Simpson, and J. A. Liggett.(2008). "Discrete blockage detection in pipelines using the frequency response diagram: Numerical study." J. Hydraul. Eng. 134 (5): 658–663. https://doi.org/10.1061/(ASCE)0733-9429(2008)134:5(658).

Lena, V.G., (2012). Pipelines Explained: How Save Are American's 2.5 Million Miles of Pipelines. 15 November. https://www.propublica.org/article/pipelines-explained-how-safe-are-americas-2.5-million-miles-of-pipelines

Li Wenqing, Gong Jing , Lü Xiaofang, Zhao Jiankui, Feng Yaorong and Yu Da. (2013) A study of hydrate plug formation in a subsea natural gas pipeline using a Novel High-pressure flow loop. Pet.Sci.(2013)10:97-105

Lijuan, W., Hongwei, Z., and Hui, J., (2012). A Leak Detection Method Based on EPANET and Genetic Algorithm in Water Distribution Systems. Software Engineering and Knowledge Engineering: Theory and Practice – Advances in Intelligent and Soft Computing 14, 459- 465.

Liou, J. (1991). "Leak detection and location by transient flow simulations." In Proc., API Pipeline Conf., 268–281. Washington, DC: American Petroleum Institute

Mashford, J., De Silva, D., Burn, S., and Marney, D., (2012). Leak Detection in simulated water pipe networks using SVM. Applied Artificial Intelligence: An International Journal 26(5), 429-444.

Mahmutoglu, Y., Turk, K., (2018). A passive acoustic based system to locate leak hole in underwater natural gas pipelines. Digit. Signal Process. 76, 59–65. https://doi.org/10.1016/j.dsp.2018.02.007.

Mitchell, T., (1997). Machine Learning. McGraw Hill, New York

Mounce, S. R., Mounce R. B., and Boxall. J. B. (2010). "Novelty detection for time series data analysis in water distribution systems using support vector machines." J. Hydroin. 13 (4): 672–686. https://doi.org/10.2166 /hydro.2010.144

Mounce S.R. and Machell, J. (2007) Burst detection using hydraulic data from water distribution systems with artificial neural networks. Urban Water Journal, 3(1):21–31, February 2007.

Moody, L.F. and Princeton N.J (1944) Friction Factors for Pipe Flow. Transactions of the American Society of Mechanical Engineers, 66, 671-681

Misiunas, D., J. Vtkovsky, G. Olsson, A. Simpson, and M. Lambert, (2005) "Pipeline break detection using pressure transient monitoring." *Journal of Water Resources Planning and Management*, **131,** 316-325 (2005).

Mountcastle V. B., (1998) Perceptual Neuroscience: The Cerebral Cortex, Harvard University Press, Cambridge, USA.

Muntakim, A., A. Dhar, and R. Dey. (2017). "Interpretation of acoustic field data for leak detection in ductile iron and copper water-distribution pipes." J. Pipeline Syst. Eng. Pract. 8 (3): 05017001. https://doi.org/10 .1061/(ASCE)PS.1949-1204.0000257

Maciej Chaczykowski, (2010).Transient flow in natural gas pipeline – The effect of pipeline thermal model, Applied Mathematical Modelling, Volume 34, Issue 4, 2010, Pages 1051-1067,ISSN 0307-904X, https://doi.org/10.1016/j.apm.2009.07.017.

Modisette, J. (2004). “Automatic tuning of pipeline models.” In Proc., PSIG Annual Meeting (SPE PSIG 0406), 1–17. Richardson, TX: Society of Petroleum Engineers.

Nasir, A., Soong, B. H., Ramachandran, S., (2010). Framework of WSN based human centric cyber physical in-pipe water monitoring system. 11th International Conference on Control, Automation, Robotics and Vision, 1257-1261.

NOSDRA. https://nosdra.oilspillmonitor.ng/ Accessed 20th November 2022

Nsofor A. Paul, Tobinson A. Briggs and Adewale Dosunmu. (2020).Technical Feasibility Study On Hydrate Transportability For Condensate Production Flow Lines. World Journal of Engineering Research and Technology. wjert, 2020, Vol. 6, Issue 2, 108-126.

P.-S. Murvay and I. Silea, (2012). “A survey on gas leak detection and localization techniques," J. Loss Prevention Process Ind., vol. 25, no. 6, pp. 966973,

Poulakis, Z., Valougeorgis, D., and Papadimitriou, C., (2003). Leakage detection in water pipe networks using a Bayesian probabilistic framework. Probabilistic Engineering Mechanics 18, 315-327.

Png, W.H., Lin, H.S., Pua, C.H., Rahman, F.A., (2018). Pipeline monitoring and leak detection using Loop integrated Mach Zehnder Interferometer optical fiber sensor.Opt. Fiber Technol. 46, 221–225. https://doi.org/10.1016/j.yofte.2018.10.013.

Prosper. G.N., Njomo. D., and Kevin Z. N., (2019) Modeling and Simulation of the Temperature Profile along Offshore Pipeline of an Oil and Gas Flow: Effect of Insulation materials. Volume 4, Issue 9, September – 2019 International Journal of Innovative Science and Research Technology

Rekha, G., Tyagi. A.K., Sreenath. N., and Mishra. S.,(2021) . "Class Imbalanced Data: Open Issues and Future Research Directions," *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 2021, pp. 1-6, doi: 10.1109/ICCCI50826.2021.9402272.

Renata T. de A. Minhoni, Francisca F. S. Pereira, Tatiane B. G. da Silva, Evanize R. Castro, João C. C. Saad.(2020). The Performance Of Explicit Formulas For Determining The Darcy Weisbach Friction Factor Engenharia Agrícola, Jaboticabal, v.40, n.2, p.258-265, mar./apr.

Richard Stover (2013) Americas Dangerous Pipelines. Center for Biological Diversity. https://www.biologicaldiversity.org/campaigns/americas_dangerous_pipelines/

Romano, M., Kapelan, Z., and Savi´c, D. A. (2010). "Real-time leak detection in water distribution systems." Proc., 12th Water Distribution Systems Analysis Conf., ASCE, Reston, VA.

Romano, M., Kapelan, Z., and Savić, D., (2011). Real-Time Leak Detection in Water Distribution Systems. Water Distribution Systems Analysis, 1074-1082.

Salam A.E.U. Tola, M. Selintung, M. and Maricar. F. (2014) On-line monitoring system of water leakage detection in pipe networks with artificial intelligence. ARPN Journal of Engineering and Applied Science, 9(10):1817–1822, October 2014.

Satelytics. (2016). Detecting Oil Pipeline Leaks Using Satellite Data: A Case Study in Eagle Ford, Texas. https://www.satelytics.com/archives/2016-detecting-oil-pipeline-leaks-using-satellite-data-a-case-study-in-eagle-ford-texas

S. Ahmad and J. Hawkins, (2017). Untangling Sequences: Behavior vs.External Causes, BioRxiv.

Saurav, S.; Malhotra, P.; TV, V.; Gugulothu, N.; Vig, L.; Agarwal, P.; and Shroff, G. (2018). Online anomaly detection with concept drift adaptation using recurrent neural networks. In Proceedings of the ACM India Joint International Conference on Data Science and Management of Data, 78–87. ACM

Scott, S.L. and Barrufet M.A.,(2003) *Worldwide Assessment of Industry Leak Detection Capabilities for Single and Multiphase Pipelines*, University of Texas, Austin (2003)

Sousa, C. A.; Miranda, S. M; Benedito, U. M. M.; Romero, O. J.(2016). Numerical simulation of transient in pipelines. Latin American Journal of Energy Research v. 3, n. 2, p. 21 – 29, 2016. DOI:10.21712/lajer.2016.v3.n2.p21-29

Shi, H., J. Gong, A. C. Zecchin, M. F. Lambert, and A. R. Simpson. 2017."Hydraulic transient wave separation algorithm using a dual-sensor with applications to pipeline condition assessment." J. Hydroinf. 19 (5): 752–765. https://doi.org/10.2166/hydro.2017.146

S. Scott and M. Barrufet, (2003) ``Worldwide assessment of industry leak detection capabilities for single & multiphase pipelines," Texas A&M Univ.,College Station, TX, USA, Project Rep., 2003.

Schlumberger (2022). PIPESIM Network Simulation / Optimization https://www.software.slb.com/products/pipesim/pipesim-networksimulation. Accessed 19th October 2022

Sukarno Pudjo, Kuntjoro Adji Sidarto, Amoranto Trisnobudi, Delint Ira Setyoadi, Nancy Rohani & Darmadi. (2007). Leak Detection Modelling and Simulation for Oil Pipeline with Artificial Intelligence Method. ITB J. Eng. Sci. Vol. 39 B, No. 1, 2007, 1-19.

Schlegl. T., Seeb¨ock. P., Sebastian. M., Waldstein. U., Schmidt-Erfurth, and Langs.G., (2017).Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In IPMI, 2017

Saranya Kunasekaran and Chellammal Suriyanarayanan (2020) Anomaly detection techniques for streaming data–An overview. Malaya Journal of Matematik, Vol. S, No. 1, 703-710, 2020 https://doi.org/10.26637/MJM0S20/0133

Sivapragasam, C., Maheswaran, R., and Venkatesh, V., (2007). ANN-based model for aiding leak detection in water distribution networks. Asian Journal of Water, Environment and Pollution 5(3), 111-114.

Smits, A.J., & Hultmark, M. (2014). Turbulence: When is the Reynolds number high enough? Access Science. Retrieved December 6, 2022, from https://doi.org/10.1036/1097-8542.YB150606. https://www.accessscience.com/content/article/aYB150606

T. R. Vrålstad, A. G. Melbye, I. M. Carlsen, and D. Llewelyn, (2013)``Comparison of leak-detection technologies for continuous monitoring of subsea production templates," SPE Projects, Facilities Construct., vol. 6, no. 2,pp. 96-103, Apr. 2013.

Taylor Lunger, Hamidreza Karami; (2019) Leak Detection in Wet Natural Gas Transportation Within Hilly Terrain Pipelines. Unconventional Resources Technology Conference Denver, Colorado, USA, 22-24 July 2019.

Liu.Y., Li. C, and Barnab´as P´oczos.(2018). Classifier two sample test for video anomaly detections. In BMVC.

Ubani, C.E., Ani, G.O., Ogbole, S.O., Okologume, W.C.(2018) Modelling Long Distance Tie-Back in Niger Delta Flow Assurance in Deep Offshore . Uniport Journal of Engineering & Scientific Research Vol. 2, Issue 1, Dec./2018.

U.S. Department of Transportation Standards, (2010) "United States Environmental Impact Statement (EIS. TransCanada's Keystone XL Project," 2010. http://www.keystonepipeline-xl.state.gov/clientsite/keystonexl.nsf/19_KXL_Sec_3.13_Risk%20Assessment.pdf?OpenFileResource

Uthman Baroudi, Anas Al-Roubaiey, and Abdullah Devendiran (2017) Pipeline Leak Detection Systems and Data Fusion: A Survey. *IEEE Access*

van Reet, J., and K. Skogman. (1987). "The effect of measurement uncertainty on real time pipeline modeling applications." In Proc., ASME Pipeline Engineering Symp. (ETCE 1987), 29–33. New York: American Society of Mechanical Engineers.

Wang, W., J. Boon, and X. Kong. (2012). "Condition assessment of live PCCP lines with free-swimming electromagnetic inspection system."In ICPTT 2012: Better pipeline infrastructure for a better life. Reston,VA: ASCE

Wang, X., Mueen, A., Ding, H. Trajcevski, G., Scheuermann, P., Keogh, E.,(2013) Experimental comparison of representation methods and distance measures for time series data. *Data Min Knowl Disc* **26**, 275–309 (2013). https://doi.org/10.1007/s10618-012-0250-5

Wang, H., Duncan, I.J.,(2014). Likelihood, causes, and consequences of focused leakage and rupture of U.S. Natural gas transmission pipelines. J. Loss Prev. Process. Ind. 30, 177–187. https://doi.org/10.1016/j.jlp.2014.05.009.

Whaley, R., R. Nicholas, and J. van Reet. (1992). “Tutorial on software-based leak detection techniques.” In Proc., PSIG Pipeline Simulation Interest Group. Richardson, TX: Society of Petroleum Engineers.

Wu, X., H. Lu, K. Huang, Z. Yuan, and X. Sun. (2015). “Mathematical model of leakage during pressure tests of oil and gas pipelines.” J. Pipeline Syst. Eng. Pract. 6 (4): 04015001. https://doi.org/10.1061/(ASCE)PS .1949-1204.0000195.

Wylie, E., V. Streeter, and L. Suo. (1993). Fluid transients in systems. Upper Saddle River, NJ: Prentice Hall.

Xia, L., Xiao-dong, W., Xin-hua, Z., Guo-jin, L., (2006). Bayesian theorem based on-line leakage detection and localization of municipal water supply network. Water and Wastewater Engineering 12.

Xinhong, L., Guoming, C., Renren, Z., Hongwei, Z., Jianmin, F.,(2018). Simulation and assessment of underwater gas release and dispersion from subsea gas pipelines leak. Journal of Process Safety and Environmental Protection 119, 46–57. https://doi.org/10.1016/j.psep.2018.07.015.

Xu, X., and B. Karney. (2017). “An overview of transient fault detection techniques.” In Modeling and monitoring of pipelines and networks: Advanced tools for automatic monitoring and supervision of pipelines, edited by C. Verde and L. Torres, 13–37. Cham, Switzerland: Springer.

Zhang Qingzhou, Zheng Yi Wu, Ming Zhao, Jingyao Qi, Yuan Huang, and Hongbin Zhao (2016) Leakage Zone Identification in Large-Scale Water Distribution Systems Using Multiclass Support Vector Machines J. Water Resour. Plann. Manage., 2016, 142(11): 04016042. ACSE.

Zheng, L., and K. Yehuda. (2013). “State of the art review of inspection technologies for condition assessment of water pipes.” Measurement 46 (1): 1–15. https://doi.org/10.1016/j.measurement.2012.05.032.

Zhou, Z., Hu, C., Xu, D., Yang, J., and Zhou, D. (2011). “Bayesian reasoning approach based recursive algorithm for online updating belief rule based expert system of pipeline leak detection.” Expert Syst. Appl., 38(4), 3937–3943.