

Detection and Localization of Leaks in Water Networks

Samer Mohammad El-Zahab

A Thesis

In the Department

of

Building, Civil and Environmental Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Building Engineering) at

Concordia University

Montreal, Quebec, Canada

January 2018

© Samer El-Zahab, 2018

CONCORDIA UNIVERSITY
School of Graduate Studies

This is to certify that the thesis prepared

By: Samer Mohammad El-Zahab

Entitled: Detection and Localization of Leaks in Water Networks

and submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (*Building Engineering*)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

_____	Chair
Dr. L. Kadem	
_____	External Examiner
Dr. B. Karney	
_____	External To Program
Dr. A. Awasthi	
_____	Examiner
Dr. O. Moselhi	
_____	Examiner
Dr. Z. Zhu	
_____	Thesis Supervisor
Dr. T. Zayed	

Approved by

Dr. F. Haghighat, Graduate Program Director

February 20, 2018

Dr. A. Asif, Dean
Faculty of Engineering and Computer Science

ABSTRACT

Detection and Localization of Leaks in Water Networks

Samer Mohammad El-Zahab, Ph.D.

Concordia University, 2018

Today, 844 million humans around the world have no access to safe drinking water. Furthermore, every 90 seconds, one child dies from water-related illnesses. Major cities lose 15% - 50% of their water and, in some cases, losses may reach up to 70%, mostly due to leaks. Therefore, it is paramount to preserve water as an invaluable resource through water networks, particularly in large cities in which leak repair may cause disruption. Municipalities usually tackle leak problems using various detection systems and technologies, often long after leaks occur; however, such efforts are not enough to detect leaks at early stages. Therefore, the main objectives of the present research are to develop and validate a leak detection system and to optimize leak repair prioritization.

The development of the leak detection models goes through several phases: (1) technology and device selection, (2) experimental work, (3) signal analysis, (4) selection of parameters, (5) machine learning model development and (6) validation of developed models. To detect leaks, vibration signals are collected through a variety of controlled experiments on PVC and ductile iron pipelines using wireless accelerometers, i.e., micro-electronic mechanical sensors (MEMS). The signals are analyzed to pinpoint leaks in water pipelines. Similarly, acoustic signals are collected from a pilot project in the city of Montreal, using noise loggers as another detection technology. The collected signals are also analyzed to detect and pinpoint the leaks. The leak detection system has presented promising results using both technologies. The developed MEMS

model is capable of accurately pinpointing leaks within 12 centimeters from the exact location. Comparatively, for noise loggers, the developed model can detect the exact leak location within a 25-cm radius for an actual leak.

The leak repair prioritization model uses two optimization techniques: (1) a well-known genetic algorithm and (2) a newly innovative Lazy Serpent Algorithm that is developed in the present research. The Lazy Serpent Algorithm has proved capable of surpassing the genetic algorithm in determining a more optimal order using much less computation time. The developed research proves that automated real-time leak detection is possible and can help governments save water resource and funds. The developed research proves the viability of accelerometers as a standalone leak detection technology and opens the door for further research and experimentations. The leak detection system model helps municipalities and water resource agencies rapidly detect leaks when they occur in real-time. The developed pinpointing models facilitate the leak repair process by precisely determine the leak location where the repair works should be conducted. The Lazy Serpent Algorithm helps municipalities better distribute their resources to maximize their desired benefits.

إهداء إلى

والدتي العزيزة "ندى بخاري" ووالدي العزيز "محمد الذهب"

أختي العزيزة "آلاء الذهب"

أساتذتي الأجلاء

وطني لبنان

شكراً لكم

لكل من يريد أن يحدث تغييراً إيجابياً في هذا العالم

DEDICATED TO

My beloved mother "Nada Boukhary" and my dear father "Mohammad El-Zahab"

My beloved sister "Alaa El-Zahab"

My esteemed mentors

My country Lebanon

Thank you

To anyone who wants to make a positive difference in this world

ACKNOWLEDGEMENT

All gratitude and praise are due to ALLAH almighty who guided and aided me towards the completion of my research work and this thesis.

I cannot express my gratitude towards my mother Nada El-Boukhary for encouraging me to pursue further education and further studies, through her sacrifices and her devotion to provide everything possible for my sister and me. I am grateful to my father Mohammad El-Zahab who tried to provide everything possible for me to grow and succeed, and who has always believed that I can do whatever I set my mind to. Much gratitude to my sister Alaa El-Zahab for her continuous encouragement and support.

I am grateful to my supervisor Prof. Tarek Zayed for giving me the opportunity to pursue a Ph.D. in Concordia University, and for supporting me and guiding me through the past three years of this thesis work, in addition to providing all the necessary tools for me to thrive and achieve. I would like to extend my thanks to the water group team leader Dr. Laya Parviz Sedghy, who always took the initiative to help all new students -myself included- succeed, and giving the necessary tips and information required to continue progressing with minimum obstacles.

I am deeply grateful to my professors at the University of Balamand and Rafik Hariri University for giving me the tools for excellence. I would like to thank Dr. Nabil Semaan who always encouraged me to pursue a Ph.D. and aided me until I received my opportunity. Many thanks go as well for Prof. Nabil Fares, whom I have learned many lessons about perfecting one's work and I hope that many of his lessons shine through this thesis. Dr. Hassan Ghanem who supervised my BSc project and helped layout my plans in academia and taught me the value of excellence and how to be a good engineer. I am honored to have been taught by unique mentors such as Dr. Najib Georges, Dr. Hikmat Zerbe, Dr. Riad Al-Wardany, Dr. Zahi Ramadan, Dr. Rached Zantout, Dr. Wajdi Al-Charif, and Ms. Nadia Jalloul, who all impacted my life in positive ways beyond imagining with their kindness and wisdom. Forgive me if I forgot anyone as the list is extensive and I only have one page.

Finally, special thanks go to my friends and family in Canada, Lebanon, and all over the world for keeping in touch with me even when I was busy and supporting me with their kind words and humor in tough times. Thank you -in alphabetical order- Abdelkader Al-Mir, Abdelrahman Dabboussi, Abdelrahman Sanjekdar, Abdelrahman Shouk, Ahmed Asaad, Alaa Mrad, Alaa Shamma, Albaraa Al-Zoabi, Eiad Alhamed, Eslam Mohammed Abdelkader, Fadi El-Darwich, Hamdi Chaouk, Hasan Ayouby, Joud El-Ali, Meriem Kabbaj, Myriam Kharma, Nour Arab, Omar Hamzeh and Tarek Al-Siddik. I am sorry if I forgot anyone under the pressure of writing.

Thank you all, and may ALLAH bless you all.

Samer El-Zahab
December 2017, Montreal, Quebec, Canada

TABLE OF CONTENTS

LIST OF FIGURES	xi
LIST OF TABLES	xvii
CHAPTER I: INTRODUCTION	1
I.1 Problem Statement	1
I.2 Research Objectives.....	2
I.3 Research Methodology Overview.....	2
I.4 Structure of the Thesis	6
CHAPTER II: LITERATURE REVIEW	8
II.1 Leak Monitoring Overview	10
II.1.1 Leak Monitoring Phases	11
II.1.2 Progress in Leak Detection Research	13
II.1.3 Leak Detection Classes.....	18
II.1.4 Leak Monitoring Models	20
II.1.5 Accelerometer Signal Analysis	32
II.2 Prioritization of Independent Events	34
II.3 Model Development Techniques.....	36
II.3.1 Artificial Neural Networks (ANN).....	36
II.3.2 Genetic Algorithms.....	39
II.3.3 Regression Analysis	41
II.3.4 Support Vector Machines (SVM)	42

II.4 Limitations of Current Approaches	44
II.4.1 Limitations of Current Technologies.....	44
II.4.2 Limitations of Prioritization Approaches	48
 CHAPTER III: ACCELEROMETER BASED MODELS	 50
III.1 Accelerometer Real-time Monitoring Model	50
III.1.1 Vibration Signal Analysis Model.....	52
III.1.2 Accelerometer Leak Detection Model.....	53
III.1.3 Accelerometer Leak Size Estimation Model.....	56
III.1.4 Accelerometer Leak Pinpointing Model	57
III.1.5 Vibration Signal Decay Model.....	58
III.2 Data Collection for Accelerometers	61
III.2.1 Experimental Work	61
III.2.2 Data collection	65
III.2.3 Developed Indicators	67
III.3 Results of Developed Models	69
III.3.1 Leak Detection and Size Identification Models.....	70
III.3.2 Accelerometer Leak Pinpointing Model Results.....	79
III.3.3 Vibration Signal Decay Model.....	99
 CHAPTER IV: NOISE LOGGERS BASED MODELS	 102
IV.1 Noise Loggers Leak Detection System	102
IV.1.1 Acoustic Signal Analysis Model	103
IV.1.2 Acoustic Leak and Pump Detection Model.....	107

IV.1.3 Acoustic Leak Pinpointing Model	111
IV.2 Data Collection for Loggers	114
IV.2.1 Zonescan System	114
IV.2.2 Data Collection	117
IV.3 Acoustic Leak Monitoring Implementation and Results	118
IV.3.1 Noise Logger Leak Identification Model.....	118
IV.3.2 Acoustic Leak Pinpointing Model Implementation and Results	127
 CHAPTER V: REPAIR PRIORITIZATION MODELS.....	 144
V.1 Repair Prioritization Model Development Methodologies	144
V.1.1 The Lazy Serpent Algorithm	144
V.1.2 Genetic Algorithm Comparative Model	164
V.2 Repair Prioritization Model Implementation	167
V.2.1 Genetic Algorithm Model Implementation and Results	167
V.2.2 Lazy Serpent Model Comparative Results.....	170
V.3 Repair Prioritization Results and Conclusions.....	174
 CHAPTER VI: DEVELOPED AUTOMATED TOOLS	 176
VI.1 Acoustic Leak Detection Tool	176
VI.1.1 Acoustic Leak Detection Tool Design.....	176
VI.1.2 Acoustic Leak Detection Tool Manual	178
VI.2 Acoustic Leak Pinpointing Tool	181
VI.2.1 Acoustic Leak Pinpointing Tool Design	182
VI.2.2 Acoustic Leak Pinpointing Tool Manual	183

VI.3 Automated Lazy Serpent Tool.....	187
VI.3.1 Automated Lazy Serpent Tool Design.....	187
VI.3.2 Automated Lazy Serpent Tool Manual.....	188
 CHAPTER VII: RESEARCH CONTRIBUTIONS AND FUTURE WORK	192
VII.1 Summary and Conclusions.....	192
VII.2 Research Contributions	194
VII.3 Research Limitations	196
VII.4 Future Work	197
VII.4.1 Enhancement Areas.....	197
VII.4.2 Extension Areas	200
 REFERENCES.....	201
 APPENDIX A: LEAK DETECTION TECHNOLOGIES.....	213
A.1 Listening Devices	213
A.2 Leak Noise Loggers.....	214
A.3 Infrared Thermography.....	214
A.4 Tracer Gas.....	215
A.5 Ground Penetrating Radar	216
A.6 Leak Detecting Robots	217
A.7 Wireless Micro-Electro-Mechanical Systems (MEMS)	221

LIST OF FIGURES

Figure I-1: General Methodology and Overview	4
Figure II-1: Literature Review Overview	9
Figure II-2: Leak Detection Literature Review Methodology	11
Figure II-3: Display of most cited papers in the leak detection field via CitNetExplorer....	14
Figure II-4: Display of most publishing countries in the leak detection field	15
Figure II-5: Top ten most utilized keywords in leak detection research by CiteSpace.....	18
Figure II-6: Schematic of buried water pipe	29
Figure II-7: Artificial Neural Network Structure.....	38
Figure II-8: Generalized Genetic Algorithm Flowchart	41
Figure III-1: Accelerometer Real-Time Monitoring System General Methodology	51
Figure III-2: Vibration Signal Analysis Model	52
Figure III-3: Leak Identification Model by Vibration Signals	54
Figure III-4: Accelerometer Leak Identification Model Development Approach.....	55
Figure III-5: Leak Size Identification Model by Vibration Signals	56
Figure III-6: Leak Pinpointing Model by Vibration Signals	57
Figure III-7: Vibration Signal Decay Model	59
Figure III-8: Decay Theory Experiment Setup	60
Figure III-9: Experimentation Setup Diagram	62
Figure III-10: Two-inch Ductile Iron Pipeline	62
Figure III-11: Pipeline Exit with Release Extension.....	63

Figure III-12: One-inch and Two-inch PVC Pipelines.....	64
Figure III-13: Time Distribution of Accelerometer Experiments	65
Figure III-14: Sample of Received Data during Experimentation.....	66
Figure III-15: Sample Matlab Analysis of Sensor Signal.....	67
Figure III-16: Sample of the Collected Data Sheet.....	69
Figure III-17: Decision Tree Leak Identification Model.....	71
Figure III-18: Naïve Bayes Leak Identification Model	72
Figure III-19: Decision Tree Leak Size Identification Model.....	75
Figure III-20: Naïve Bayes Leak Size Identification Model	76
Figure III-21: Box and Whisker Plot for Leak Identification Models.....	78
Figure III-22: General X_L Regression Prediction Report	81
Figure III-23: General X_R Regression Prediction Report.....	82
Figure III-24: X_L for Ductile Iron Regression Prediction Report.....	83
Figure III-25: X_R for Ductile Iron Regression Prediction Report.....	84
Figure III-26: X_L for PVC Regression Prediction Report	86
Figure III-27: X_R for PVC General Prediction Report	87
Figure III-28: Sensitivity Analysis of the General X_L and X_R Models.....	90
Figure III-29: Sensitivity Analysis of the X_L and X_R Models for Ductile Iron.....	91
Figure III-30: Sensitivity Analysis of the X_L and X_R Models for PVC Pipelines.....	92
Figure III-31: Deviation Data Distribution for General X_L and X_R Equations	97
Figure III-32: Deviation Data Distribution for Ductile Iron X_L and X_R Equations	98
Figure III-33: Deviation Data Distribution for PVC X_L and X_R Equations.....	98

Figure IV-1: Acoustic Real-Time Monitoring System Model.....	102
Figure IV-2: Frequency Distribution Analysis.....	103
Figure IV-3: Acoustic Signal Analysis Model	104
Figure IV-4: Identification by Aggregation Model.....	108
Figure IV-5: Aggregator Methodology	111
Figure IV-6: Acoustic Pinpointing Model Methodology	113
Figure IV-7: Zonescan Maintenance Interface	115
Figure IV-8: Zonescan Leak Pinpointing Interface.....	116
Figure IV-9: Zonescan Leak Cross Correlation Report	117
Figure IV-10: Decision Tree Model for Identifying Logger Leaks	119
Figure IV-11: Accuracy of Decision Tree Leak Identification Model for Noise Loggers ..	120
Figure IV-12: Accuracy of Naïve Bayes Leak Identification Model for Noise Loggers	120
Figure IV-13: Accuracy of Deep Learning Leak Identification Model for Noise Loggers	121
Figure IV-14: Accuracy Change Impact for One Model With Two Models at 87%	121
Figure IV-15: α_{AMP} for One Changing Model with Two Models at 87%	122
Figure IV-16: Impact Envelope for Two Models Aggregated with an 89% Model.....	123
Figure IV-17: α_{AMP} for Two Models Aggregated With One Model at 89%	124
Figure IV-18: Aggregation Impact of Three Changing Models	125
Figure IV-19: α_{AMP} of Three Models Changing Accuracies	126
Figure IV-20: Best Subset Analysis Report for X_L	128
Figure IV-21: Acoustic $X_L(1)$ Regression Prediction Report.....	130
Figure IV-22: Acoustic $X_R(1)$ Regression Prediction Report.....	131
Figure IV-23: Acoustic $X_L(2)$ Regression Prediction Report.....	133

Figure IV-24: Acoustic $X_R(2)$ Regression Prediction Report.....	134
Figure IV-25: Acoustic X_{LF} Regression Prediction Report.....	136
Figure IV-26: Acoustic X_{RF} Regression Prediction Report.....	137
Figure IV-27: Acoustic PX_L Regression Prediction Report.....	138
Figure IV-28: Acoustic PX_R Regression Prediction Report.....	139
Figure V-1: Lazy Serpent Algorithm Preliminary Model.....	145
Figure V-2: Lazy Serpent Algorithm Components Summary.....	146
Figure V-3: The Reaction of Serpents to Basic Events.....	150
Figure V-4: The Reaction of Serpents to Serpent Specific Events	151
Figure V-5: The Reaction of Serpents to Simultaneous Multi Serpent Events.....	152
Figure V-6: The Reaction of Serpents to a Consecutive Layers Multi Serpent Events	154
Figure V-7: The Reaction of Serpents to Any Serpent Events	155
Figure V-8: Returning Serpents Schematic Explanation.....	156
Figure V-9: Independent Serpent Behavior	157
Figure V-10: Pack Mode Serpents Behavior	158
Figure V-11: Inverse Pyramid Solution Approach.....	163
Figure V-12: Updated Budget Equation Methodology	164
Figure V-13: Genetic Algorithm Model Methodology	165
Figure V-14: Genetic Algorithm Input Excel Sheet	167
Figure V-15: Genetic Algorithm Progress through Trials.....	168
Figure V-16: Optimal Order of Fictitious Example	169
Figure V-17: Distribution of Benefit-Cost Ratio Condition at Repair Time.....	169

Figure V-18: Lazy Serpent Software Interface Input	171
Figure V-19: Basic Lazy Serpent Repair Order	172
Figure V-20: Inverse Lazy Serpent Repair Order.....	173
Figure V-21: Selective Lazy Serpent Proposed Repair Order	174
Figure VI-1: Leak Detection Tool Processing Flowchart.....	177
Figure VI-2: Pump and Leak Identifier Interface	178
Figure VI-3: Software Browsing Interface	179
Figure VI-4: Wave Representation of an Analyzed File	180
Figure VI-5: Results of the Analysis.....	181
Figure VI-6: Leak Pinpointing Tool Design Flowchart.....	182
Figure VI-7: Initial Leak Pinpointing Interface	183
Figure VI-8: Adding The Value of Distance D.....	184
Figure VI-9: Pinpointing Tool Browsing Panel.....	185
Figure VI-10: Pinpointing Signal Analysis Results.....	186
Figure VI-11: Leak Pinpointing Result Interface	186
Figure VI-12: Automated Lazy Serpent Design Flowchart	187
Figure VI-13: Automated Lazy Serpent Initial Interface	188
Figure VI-14: Automated Lazy Serpent Constraint Definition.....	189
Figure VI-15: Automated Lazy Serpent Data Input	190
Figure VI-16: Automated Lazy Serpent Result Output.....	191
Figure A-1: Infrared image of a floor	215

Figure A-2: GPR Leak detection of a simulated leak at multiple angles.....	216
Figure A-3: Movement of Smart-Ball within a pipeline with possible results.	218
Figure A-4: Overall Shape and architecture of Explorer leak detection robot.	219
Figure A-5: Beaver robot along with the van that operates and displays the robot.....	220
Figure A-6: Pear-point leak detection system	220

LIST OF TABLES

Table II-1: Phases of Leak Detection - ILLP.....	13
Table II-2: Comparison of the Available Technologies against Multiple Criteria.	48
Table III-1: Leak Detection Model Cross Validation Results	73
Table III-2: Leak Size Identification Model Cross Validation Results.....	77
Table III-3: Leak Identification Model Thresholds.....	79
Table III-4: Statistical Results of the Developed ANN Models.....	88
Table III-5: Statistical Analysis of the Developed Regression Models.....	93
Table III-6: Advanced Statistical Analysis of the Developed Regression Models	94
Table III-7: Validation Results of Developed General Model.	96
Table III-8: Vibration Signal Decay Model Results	99
Table IV-1: Summary of Collected Acoustic Data.....	118
Table IV-2: Aggregation Results of Leak Identification Model	127
Table IV-3: Advanced Statistical Analysis of Acoustic Leak Pinpointing Models	141
Table IV-4: Viger Street Leak Validation of Leak Pinpointing Models.....	143
Table V-1: Comparison of Prioritization Models	175

CHAPTER I: INTRODUCTION

I.1 Problem Statement

Nearly one billion people worldwide do not have access to clean drinking water (Kretnak 2016). Substantial quantities of drinking water are lost in water distribution systems mainly due to leaks, which often leads to potential water losses. The percentage of water loss is around 20% - 30% of the treated water for urban consumption in distribution networks mainly due to leaks (Cheong 1991), and in some systems, this loss can surpass 50% of the produced water (AWWA 1987). The contribution of leaks to the total water loss within distribution networks is estimated to be at 70% and is expected to rise in low maintenance locations (Van Zyl and Clayton 2007). Water loss is not the only outcome of leaks, as they also create problems at the social and environmental levels. For example, the United Kingdom is estimated to dig approximately 4 million holes annually into its road network to install pipes and repair water leaks. The overall cost of the damages created by leaks in the UK is estimated to be GPD 7 billion a year (around USD 10 billion) divided into GPD 1.5 billion (around USD 2.16 billion) of direct damage costs and GPD 5.5 billion (around USD 8 billion) of social impact costs (Royal et al. 2011). Furthermore, leaks that are left unrepaired are susceptible to grow and thus allow pathogens and contaminants from the environment into the water network, which results in a significant decrease in the quality of the provided water and might harmfully affect the lives of humans and other living species (Alkassseh et al. 2013).

In response to these damages and negative impacts, researchers have developed a real-time monitoring system within water distribution networks that allows early detection of leaks and eventually optimally-timed repairs. Multiple models were developed to address the issue. Nevertheless, some have had limitations in terms of accuracy, device availability, applicability,

false alarms and the impacts of external conditions. Hence, this research proposes a real-time monitoring system for pressurized water networks, capable of detecting, localizing and pinpointing leaks by utilizing accelerometers that are widely available in the market.

I.2 Research Objectives

The primary objective of this research is to develop an automated wireless real-time monitoring system for pressurized networks, which is coupled with an algorithm capable of prioritizing leak repairs as a function of the priority of the damaged pipeline. To fulfill this primary objective, the following sub-objectives are developed:

1. Identify, study, and select leak detection technologies.
2. Develop and validate technology-based leak detection and pinpointing models.
3. Build an optimized model for leak repair prioritization.
4. Automate the developed models.

I.3 Research Methodology Overview

This research proposes a detailed framework for a real-time monitoring system for pressurized water networks and a leak prioritization algorithm. The overall flow diagram of this research is displayed in Figure I-1. The figure highlights three main topics as follows: (1) Literature review, (2) model development and (3) model verification and experimentation. Figure I-1 illustrates the reviewed literature in four main categories: (1) Leak detection models and approaches, (2) Leak detection technologies, (3) Prioritization algorithms and (4) Simulation techniques. Leak detection approaches are techniques that rely mainly on a mathematical model that interprets the readings of a leak detecting technology, whereas leak detection technologies are modern day

technologies that have proven to be capable of reacting to the existence of leaks. The two fields are interconnected and interrelated and thus difficult to be separated. In this research, the aim is to identify technologies that are capable of detecting the existence of leaks in real-time and to review how previous research have tackled the problem of interpreting the data derived from the available technologies. Prioritization algorithms are studied to identify what algorithms are currently used in the industry and what is missing in those algorithms. The limitations help establish the enhancements provided by the newly proposed algorithm, i.e. the lazy serpent, compared to the currently used algorithms. The study of simulation techniques is essential to the newly developed algorithm as one of the lazy serpent's contributions and the advantages are simulating the progression of the leaks and identifying the constant change in state. Completing the aforementioned literature review completes the first objective of this thesis mentioned in section I.1 and that is to identify, study, and select leak detection technologies.

To complete the second objective, two sets of leak detection models are developed. The technologies are Accelerometers and Noise Loggers. The models are described in Chapters 3 and Chapter 4 respectively. As far as the models are concerned, the first principal model is the real-time monitoring system. The goal of the pressurized water network monitoring system is to identify the existence of leaks in the first place and then pinpoint the leak. The model also explores possible approaches to identifying the size of the leak and how the vibration signal moves in the pipeline. In terms of model verification, multiple experiments are done on one-inch and two-inch pipelines made of PVC and ductile iron to develop and verify the leak detection model. The experiments are designed to identify the behavior of leaks and their induced vibration signal as well as to develop the models above and then verify these models.

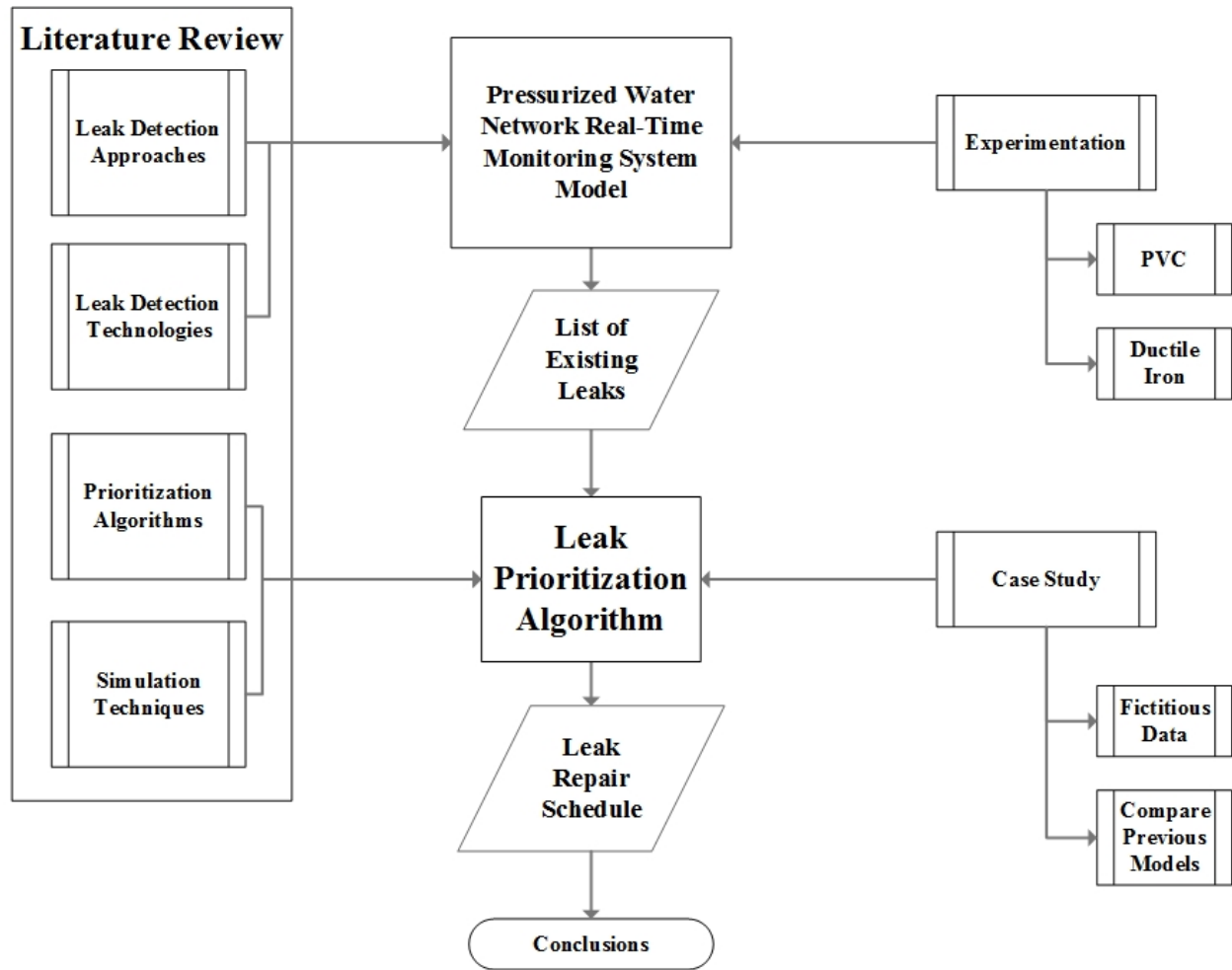


Figure I-1: General Methodology and Overview

The third objective is tackled by the development of two repair prioritization models. The models are described in Chapter 5. The chapter introduces the algorithm named “Lazy Serpent.” Lazy Serpent aims to simulate the deterioration of each leak within a three-dimensional landscape where the axis x, y, and z represent three main factors that govern the deterioration and the optimization factors that are set by the user. When it comes to the Lazy Serpent Algorithm, case studies from previous research are simulated in this algorithm and their output

are compared to previous results. Fictitious data will be also developed and simulated to discover the extent and capabilities of the newly developed algorithm.

Finally, the fourth objective of this research is fulfilled in Chapter 6. Chapter 6 introduces three computer-based tools. The tools are for leak identification, leak pinpointing, and an automated version of the lazy serpent. In summary, the overall research methodology can be described as follows:

Step 1: Literature Review

A comprehensive state-of-the-art literature review presents leak detection technologies and leak detection models. A leak detection technology is a technology – hardware – that is capable of reacting to the existence of leaks. On the other hand, leak detection models are mathematical and analytical approaches that utilize technology readings to detect and identify leaks. This literature review is intended to be expanded by an overview of the-state-of-the-art prioritization algorithms.

Step 2: Signal Analysis Model

After the selection of accelerometers and noise loggers as the technologies this research develops, the need for deciphering the readings from the devices was identified. Thus, following several trials, a model was adopted based on the literature review to overcome the problem of signal deciphering and analysis.

Step 3: Leak Detection and Localization Model

This model presents a new way to pinpoint leaks by utilizing the readings from accelerometers and noise loggers to identify, localize, and pinpoint leaks. The presented models are based on experiments in addition to statistical analysis techniques, such as regression analysis and artificial neural networks.

Step 4: Leak Repair Prioritization Model

In this step, a new three-dimensional prioritization algorithm is developed using a combination of simulation techniques and optimization. This algorithm aims to minimize the damage of leaks by setting an optimal leak repair prioritization approach.

Step 5: Software Development

Finally, a software is developed to automate the whole framework. This software is expected to be user-friendly and useful to municipalities.

I.4 Structure of the Thesis

This thesis is composed of seven chapters summarized as follows:

Chapter 1 – Introduction

This chapter sheds light on the research problem at hand, introduces the thesis contents and summarizes the methodology.

Chapter 2 – Literature Review

The chapter reviews previous leak detection models as well as available leak detection technologies. Furthermore, previous approaches and techniques are compared to establish the advantages and the limitations of the previous models and available technologies.

Chapter 3 – Accelerometer Based Models

The chapter presents a detailed description of the methodology utilized for the development of the models using accelerometers. The chapter also describes the experiments conducted to collect data from the devices, the models developed, and the results achieved by the models.

Chapter 4 – Noise Logger Based Models

The chapter describes the methodology for noise loggers followed by the implementation of the methodology and data collection that were conducted with the city of Montreal. The developed models are presented, and their results are discussed.

Chapter 5 – Repair Prioritization Models

The chapter presents the methodology used for the development of two approaches for leak repair prioritization, the genetic algorithm approach and the lazy serpent. The implementation of the two models is presented and their results are compared and discussed.

Chapter 6 – Developed Automated Tools

The chapter discusses three automated tools that are developed to accompany the model. Chapter 6 goes over the structure and design of the developed tools and presents a user walkthrough or manual for the user of the tools.

Chapter 7 – Research Contributions and Future Work

The chapter concludes this research by presenting its current contributions, limitations and the expected future work. Chapter 7 also discussed the possible future advancements.

CHAPTER II: LITERATURE REVIEW

This chapter summarizes the previous research in leak detection and independent event prioritization. Figure II-1 illustrates the steps of conducting the literature review. The literature review is composed of two main categories: (1) Leak detection literature and (2) Independent event scheduling and prioritization literature. The bulk of the literature review is on leak detection, covering the topics of leak detection phases, current leak detection models, available leak detection technologies and model development techniques. Meanwhile, accelerometer signal analysis is presented separately due to its importance in comprehending the derived data from the selected devices although it is basically relevant to the leak detection technologies section. This chapter serves to accomplish the first objective of this thesis that is described in Chapter 1 and the objective aims to identify, study, and select leak detection technologies.

As its primary outcome, research on the leak detection phases helps identify the general approaches and steps, to detect the exact leak location. In terms of leak detection models, the aim of the literature review is to review available models and their techniques and move on to identify the limitations of those models. Additionally, leak detection technologies are addressed and compared, to identify the current state of the art and their limitations. Model development techniques in leak detection were explored to determine suitable techniques for this research. Finally, a signal analysis model for vibration signals was adopted. Independent event prioritization was adopted mainly because leaks are assumed to be independent of one another and hence conventional scheduling techniques are not viable; thus, another approach must be utilized. In this part, independent event prioritization approaches are presented along with their limitations.

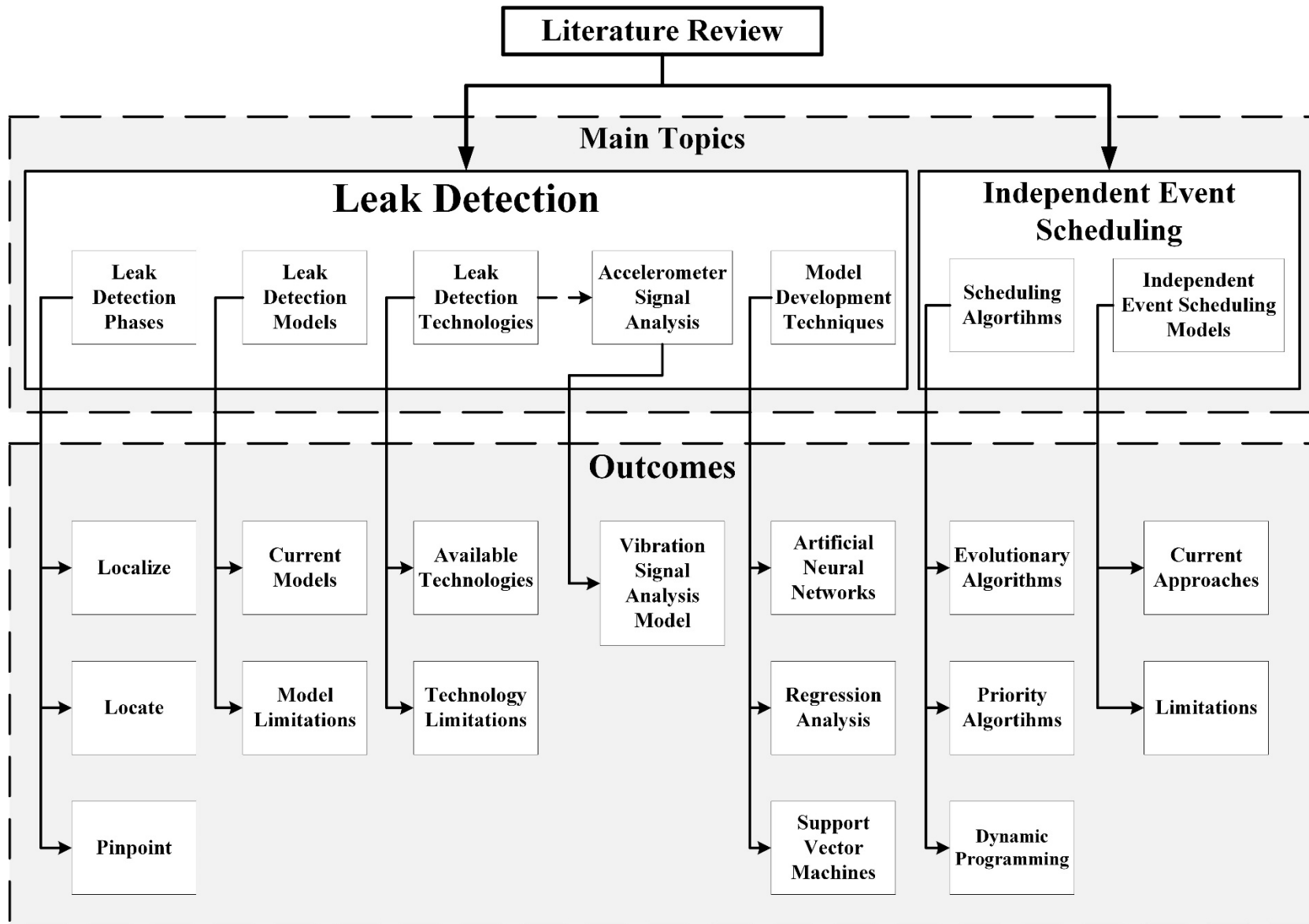


Figure II-1: Literature Review Overview

II.1 Leak Monitoring Overview

This section provides an overview of the current state-of-the-art approaches and understandings in leak detection. Here, the phases of leak monitoring are identified and the difference between their definitions are explained. The state-of-the-art models and technologies are then discussed. Finally, a model for the analysis of vibrational signals is presented. The overall methodology of this literature survey can be summarized in Figure II-2. The first step of this research is to gather a substantial amount of the published works on leak detection in urban water mains. For this purpose, the Web of Knowledge (WoK) database was selected as the main database to conduct the search. The main keywords of the search are “leak detection in pipelines”. The data collected from the WoK database will be analyzed to have a sense of the historical, technological, and demographical trends in pipeline leak detection research. The second step is to randomly select between 30 to 35 papers on the topic and analyze the chosen papers with more depth. This approach creates two main datasets, the WoK dataset which contains nearly 1000 papers on the topic of leak detection pipelines and the in-depth dataset for this research article that studies a randomly selected collection of papers. The in-depth dataset will be analyzed regarding capabilities and technologies used. Additionally, the papers will be assessed for their historical progression and the distribution of the models regarding technologies used and capabilities provided. This analysis allows the development of a general sense of the progression and flow of leak detection research.

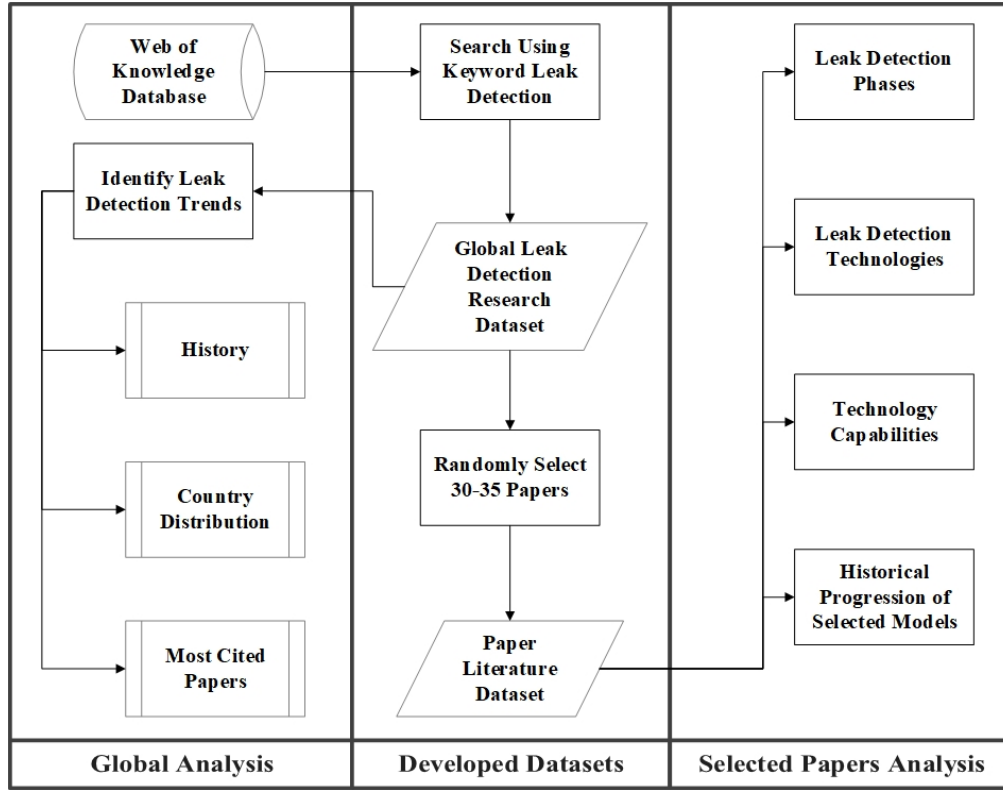


Figure II-2: Leak Detection Literature Review Methodology

II.1.1 Leak Monitoring Phases

In 2009, Hamilton defined leak detection as a subsidiary of three main phases known as localize, locate, and pinpoint or LLP. In Table II-1, localize is defined as narrowing down a leak to a specific segment within the network or a specific district metered area after the suspicion of a leak. Several tools can be used in this domain such as district metered area and fitting surveys (Li et al. 2011). The second phase is locating, and it is the first phase where a location for the leak is determined. The location of the leak is said to be within a radius of 30 centimeters. Furthermore, multiple tools are available for utility in this phase including correlators and microelectromechanical sensors (El-abbasy et al. 2014; El-Zahab et al. 2016; Martini et al. 2015). Finally, the third phase, according to Hamilton, would be pinpointing. Pinpointing is

another phase where another estimate for the exact location of the leak is determined but within a radius of 20 centimeters. In this phase, a variety of tools are available including hydrophones and geophones (Fantozzi et al. 2009; Royal et al. 2011). In contrast, multiple researchers in the field of leak detection have highlighted the existence of another challenging aspect in the development of leak detection systems and approaches. The challenge lies in differentiating the signs of leaks from other aspects such as pumps or an open hydrant. This problem is vivid in acoustic noise loggers and accelerometers and other wireless sensor devices that are used in leak detection. Sensors would pick up any form of signs and signals similar to those of leaks and end up providing false alarms. False alarms create an expenditure of workforce and funds for the bodies monitoring the respective networks (El-Zahab et al. 2016; Khulief et al. 2012; Stoianov et al. 2007a). Therefore, a new phase of leak detection is proposed, and that is the identification phase. As described in Table II-1, the identification phase works towards determining if the signs detect and the signals derived indicate a leak in the network of pipelines or not and how to differentiate between leaks and other factors affecting the network. Subsequently, the leak detection phases can now be summed up as ILLP, identify-localize-locate-pinpoint. It is highly possible to merge locating and pinpointing due to the 10-centimeter difference between the two phases. This approach helps in creating three distinct and unambiguous phases. Thus, another approach would be the ILP approach, identification-localization-pinpointing. Where the first step identifies the existence of a leak, the second phase identifies the segment where the leak is, and finally the third phase would determine the exact location of the leak with a certain accuracy.

Table II-1: Phases of Leak Detection - ILLP

Phase	Definition
Identify	Determine the existence of a leak and distinguish leaks from false alarms.
Localize	Narrow down the location of a leak to a specific area or segment.
Locate	Determine the location of a leak with a radius of 30cm.
Pinpoint	Determine the location of a leak with a radius of 20cm.

II.1.2 Progress in Leak Detection Research

The advances in leak detection research have grown remarkably in recent years, but the field of leak detection is not a new field of research. The damages created by leaks in pipeline networks presented researches with the curiosity of developing reliable and immediate solutions to fight the leakage epidemic. In a survey performed on the Web of Knowledge database, 941 scholarly articles about the topic of leak detection in pipelines were found. A timeline analysis was conducted for the most referenced papers in the field of leak detection using the CitNetExplorer software for bibliometric analysis (Van Eck and Waltman 2014). The results of this analysis are illustrated in Figure II-3, The first paper found by this survey existed in the year 1968 by Zielke as displayed in Figure II-3, who suggested the study of wall shear stress in laminar pipe flow along with the mean velocity of the flow and the changes in the velocity can aid in detecting any anomalies within the pipeline. The figure further displays that the research in the field of leak detection started booming further in the mid-1980s. Beyond that point, the research field started gaining more and more attention especially after 1994. After the year 2000, research in leak detection saw numerous publications and contributions with the rise of the importance of water conservation and water scarcity and the abundance of novel technologies that are capable of facilitating and automating the leak detection process.

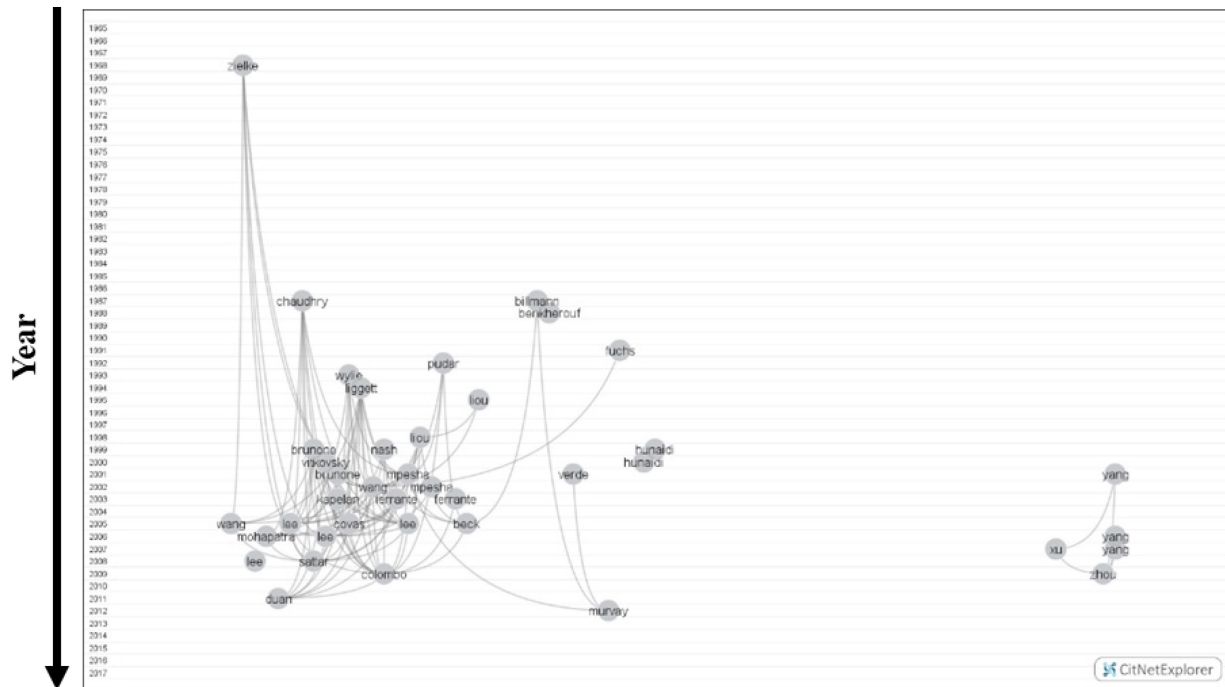


Figure II-3: Display of most cited papers in the leak detection field via CitNetExplorer

The dataset was further studied using another bibliometric analysis software that is named VOSviewer (Van Eck and Waltman 2010). Figure II-4 was established using the software. The figure aims to establish the countries with the most interest in researching the field of pipeline leak detection and to provide a sense of the number of publications provided by those countries. In the figure, the countries with at least 25 publications in the field of pipeline leak detection are displayed. Starting with Germany with 25 publications in the field within the Web of Knowledge database and moving all the way up to Canada with 58 publications in the field, then England with 74 publications in the field, exceeded by the United States of America with a sum of 144 publications, and finally on top of the list is the People's Republic of China with 263 publications. The proximity between the globes highlights the amount of co-authored works as well. Therefore, from the figure, it is deductible that the People's Republic of China had multiple

co-authored works with England. Same goes for Australia and South Korea as well as the United States of America, Germany, and Saudi Arabia.

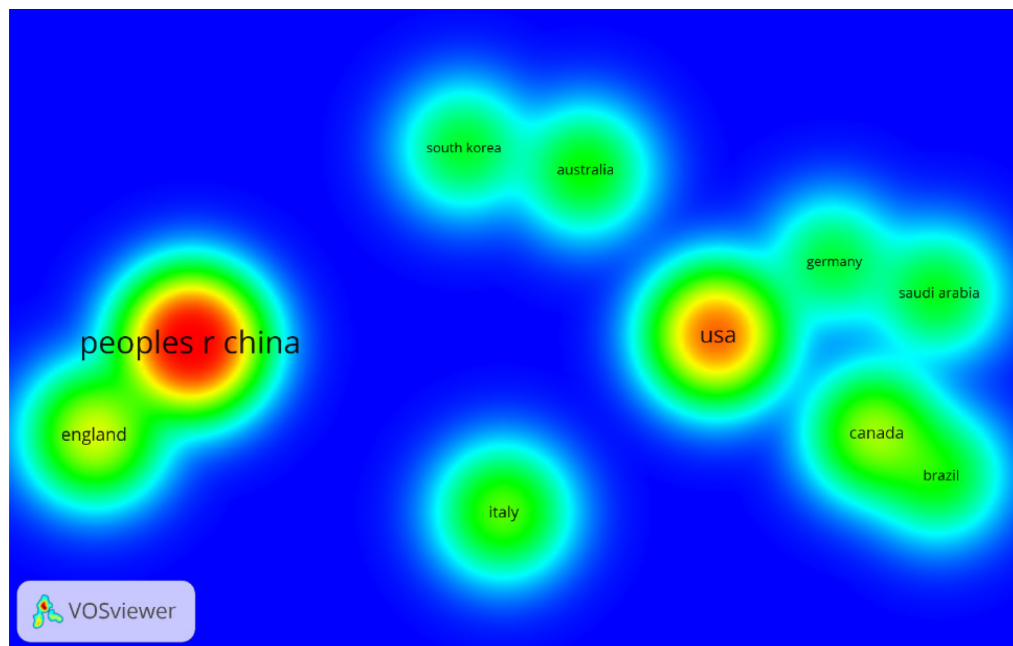


Figure II-4: Display of most publishing countries in the leak detection field via VOSviewer

After studying the geographical distribution of research interests, the next step is to analyze the most repetitive keywords in the global dataset provided by WoK. Figure II-5 displays the top ten most used words within the collected dataset. The figure starts at number 0 and ends at number 11 skipping the values 6 and 10 due to a software glitch. The first keyword with the most prominent recurrence is “gas pipelines”. The importance of gas pipelines is relative to the criticalness of the material they transport. Natural gas and oil are hazardous materials that are transported in a pressurized state. Therefore, any leakage will have a tremendous impact in a small time. The impacts of leaks in gas pipelines include major economic losses along with the pollution of the surrounding environment and a grand possibility for injuring personnel (Geiger et al. 2006; Inaudi et al. 2008; Sun et al. 2011). The second most common keyword is “short

period analysis”. The keyword refers to the analysis that focus on the first characteristic time of the pipe. The analysis revolves mainly around pressure transients within pipelines due to leaks and conducts analysis based on the time and frequency domains and it considers the conditions at minor losses as boundaries for the analysis (Meniconi et al. 2010). The third keyword is “forecasting model” which are also known as prediction models. Prediction models play key roles in leak detection as they utilize learned historical data to make calculated decision for future leaks. Additionally, forecasting models can learn and progress through time as they incur new information. Forecasting models are usually used with leak detection technologies to develop thresholds and equations that are capable of detecting and locating leaks (El-Abbasy et al. 2016; El-Zahab et al. 2016; Fahmy and Moselhi 2009; Whittle et al. 2010). The fourth most recurring keyword in the dataset is “energy analysis”. The term energy analysis refers to the study of the anomalies created by the reflected energy caused by leaks in pipelines. The term may also refer to the study of the variation of energy consumption by pumps surrounding leaks to maintain a stable pressure. In both cases the persistence of a disruption indicates the existence of leakage and accordingly it can be localized (Belouchrani et al. 2013; Mostafapour and Davoudi 2013; Shibley 2013). The fifth most recurring keyword is “frequency response diagram”. The figure shows that the frequency response analysis is the oldest recurring keyword within the data set under study. Frequency response is one of the oldest techniques in leak detection. It relies on studying the response to an excitation frequency within a pipeline. The method studies the frequency diagram provided as a result of the excitation and accordingly detects leaks (Brennan et al. 2007; Mpesha et al. 2001). The keyword with the sixth place in recurrences is “leak diagnosis algorithm”. Leak diagnosis algorithms are mathematical or artificial intelligence and machine learning models that utilize available data to automatically

identify, locate, or pinpoint leaks. Those algorithms can be paired with a wide variety of technologies such as infrared, ground penetrating radar, and noise loggers. Some of the algorithms that are used include support vector machines, artificial neural networks, and genetic algorithms (El-Abbasy et al. 2016; El-Zahab et al. 2017; Fahmy and Moselhi 2009; Al Hawari et al. 2015, 2016). In the seventh place comes the keyword “zigzag pipeline” which refers to pipelines that are not linear and have some turns and unique shapes within their structure. Those pipelines include pipelines that are part of an L-shape or T-shape structure. The importance of those pipelines lies in the fact that the unique changes in their shapes create new challenges and parameters for leak detection through altering the collected data. Therefore, extra care and attention is invested in this abundant class of pipelines (Datta and Sarkar 2016; Lay-Ekuakille et al. 2009, 2010). In the eighth position of the list comes the keyword “pipe diameter”. The diameters of pipelines that are studied for leaks have proven to be a critical factor in the leak detection process for multiple technologies and techniques. The smaller a pipeline is, the more likely it is for the leak signal to travel longer distances based on the leak size. Therefore, the diameter of pipelines is an important parameter in multiple research works (Covas et al. 2005; Hauge et al. 2007). In the ninth standing is the keyword “underground pipe”. Underground pipelines are a crucial element in the leak detection study as urban infrastructure systems are mostly deployed underground. That is why a substantial amount of research is conducted on underground pipes and also simulative experiments where pipes are tested in underground-like conditions (Mashford et al. 2009; Rajani and Kleiner 2001; Stoianov et al. 2007b). Finally, in the tenth position is the keyword “in-pipe system”, which is a form of a static leak detection system. In-pipe systems rely on the placement of sensors within networks. Those sensors are connected to a main data collection server by means of a communication technology such as 3G. The data

is then received in a timely fashion at the headquarters of the operator. Accordingly, analysis can be conducted in a timely fashion as well or using software. In-pipe systems allow for the immediate detection of leaks and that is why they are a growing leak detection topic recently (Stoianov et al. 2007; Srirangarajan et al. 2013; El-Abbasy et al. 2016; El-Zahab et al. 2017).

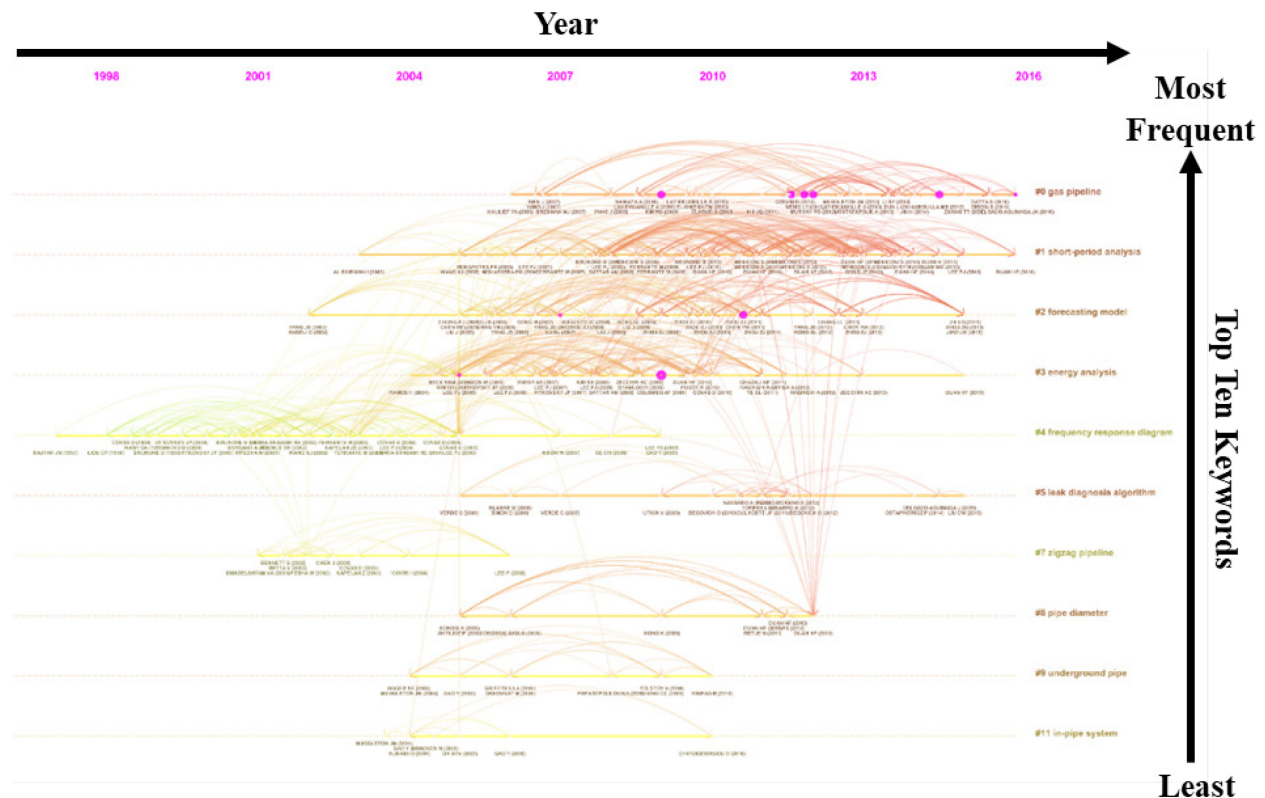


Figure II-5: Top ten most utilized keywords in leak detection research by CiteSpace

II.1.3 Leak Detection Classes

Through the observation of the literature and the applied works in leak detection, two main categories of leak detection systems can be identified. The categories are static (or stationary) leak detection and dynamic (or mobile) leak detection. Although each class on its own is capable of identifying, locating, and pinpointing leaks, it is not uncommon to utilize a

combination of both classes (Atef et al. 2016; Billmann and Isermann 1987; Romano et al. 2017). The two classes of leak detection system can be defined as follows:

- ***Static leak detection systems***: are systems that rely on sensors and data collectors that are placed within the water network and on valves and are capable of transmitting periodical data to the network management office. This data can be used to identify, localize, and pinpoint leaks.
- ***Dynamic leak detection systems***: are systems that rely on moving leak detection devices to suspected leakage area to perform an investigation. Therefore, they rely initially on suspicion of an existing leak. Another approach is performing regular surveys around cities to identify leaks as soon as possible. Those systems can confirm the existence of leaks and immediately localize and pinpoint them.

The main distinction between the two classes is that static leak detection systems can inform the water network management of the existence of a leak almost immediately, whereas dynamic leak detection systems are required to have information of a leak possibility so that they can be mobilized for investigation. On the other hand, dynamic leak detection systems can pinpoint the exact location of a leak almost immediately under ideal operating conditions, whereas static leak detection systems will provide a location within a certain area and they are also more prone to false alarms. It is not uncommon to use a static leak detection system to detect leak and a dynamic leak detection system to pinpoint them, but that is not expected to be the most affordable route (Cataldo et al. 2014; Lee et al. 2005). The two classes encompass a wide variety of technologies to provide an accurate leak detection system, but the technologies are not limited to one class. For example, acoustic technologies, specifically noise loggers, can be dynamic and

moved from one location to the other periodically to detect leaks as in (Hunaidi and Wang 2006) or they can be left in the network as in (El-Zahab et al. 2017).

II.1.4 Leak Monitoring Models

Researchers have kept a keen eye on the problem of leakage and water loss in water networks, for minimizing the corresponding losses. Thus, multiple models were developed, starting with mathematical models and going through pressure transducers and infrared models all the way to custom made devices for detecting water leaks. Geiger (2006) classified leak detection systems into two main categories: (1) Externally based systems and (2) Internally based systems. Externally based systems rely on sensors that read the information provided by the pipeline and then process the information to detect leaks and identify leak locations, whereas internally based systems rely on measurements of flow and network data to detect the leaks. Externally based systems include acoustic detectors, vibration sensors, fiber optic sensing cables, vapor or liquid sensing tubes and liquid sensing cables. Internally based systems include a wide range of mathematical models, such as line balance and volume balance as well as transient models, statistical analysis models and pressure and flow monitoring.

Mathematical Models

Mathematical models are the earliest techniques for leak detection; they relied on basic input measurement and output estimation through mathematical equations, to detect the leak and identify its possible location. Mathematical models are still widely used because their cost is minimal compared to other approaches. Mathematical models rely on simple readings regarding the condition on the water network, such as pressure and flow rate. The most commonly utilized

approaches are Mass/Volume Balance and Transient Analysis. For Mass/Volume Balance, the calculations are based on pressure and flow measurements at the inlets and outlets of pipelines, referred to as ‘nodes’. Based on the collected data, the network is divided into orifice areas and each area serves as an unknown variable. Through utilizing the collected data, the network would be converted to an inverse problem that is solved using a specific set of equations. The authors identified three types of mathematical representations to tackle: (1) Underdetermined problems, as problems with less data points than the expected number of leaks, (2) overdetermined problems, with an amount of data larger than that of the number of expected leaks – this type of problems was the most favorable, and (3) undetermined problems, which have been analyzed multiple times, yet showing inconclusive results. The output of this model identifies the segment within the network that is most likely affected by a leak (Pudar and Liggett 1992). The limitations of the aforementioned approach are: (1) the need for a significant amount of data in order to be accurate, (2) inaccuracy in identifying the location of the leak, i.e. the segment could span multiple kilometers and (3) the possibility of giving incorrect or inaccurate results due to the low quantity or precision of the data collected. Those limitations can be accounted for through utilizing transient analysis within the technique, accompanied with a significant amount of data to have a more accurate estimate of the friction factor. To increase the accuracy of the model, it is critical that the measurements of pressure and flow be taken at vulnerable and sensitive locations (Liggett and Chen 1994).

On the other hand, transient analysis is only capable of detecting leaks using the pressure and flow rate data within the network. Transient analysis can detect leaks in water mains and gas transmission networks. This model utilizes pipe flow rate values into a set of two coupled one-dimensional first order nonlinear hyperbolic partial differential equations to determine the state

of the flow and the existence of leaks. The system consists of an adaptive Luenberger-type observer and a set heuristic update laws. In terms of its limitations, the model requires improvements for more accuracy and the utilized technologies should be enhanced (Hauge et al. 2007).

Transient analysis has been improved in terms of time-frequency analysis. Array processing along with time-frequency distributions and quadratic time-frequency distributions make it possible to explore various aspects of leak signals. This approach provides a significantly more accurate calculation for the identification of leaks and origin signal recovery (Belouchrani et al. 2013). The transient analysis required further development to be capable of detecting multiple leaks at the same time. Comparing changes between the leak state and the steady state, i.e. preceding a leak occurrence, reduces the number of unknown parameters in identifying leaks and eventually minimizing the quadratic error within the calculation. This approach was tested with 12 scenarios over a 200-meter pipeline and showed results with less than 1% of deviation from the correct value (Verde et al. 2014).

With the current technological advancement, new mathematical models continue to rise in the field of leak detection. The mathematical model presented by Piller and Van Zyl (2014) is one example of an approach for detecting leaks mathematically. The authors presented a method to solve the Fixed And Varied Area Discharges (FAVAD), modeling the hydraulic equations using energy minimization equations. Here, a damped Newton algorithm is used as a tool for solving the FAVAD equations.

Pressure Transducers

Pressure transducers are a conventional approach intuitively using pressure measurements to detect and locate leaks. Similar to other techniques, pressure transducers need improvements to match the needs of modern-day technologies. This approach can be improved by an online computational approach to analyze the hydraulic transients within a single pipeline as a result of leaks (Silva et al. 1996). A lot of experiments were done in the aforementioned research work, using two PVC pipelines of 3/4-inch diameter pipelines. The two pipelines were of 433 meters and 1.248 kilometers. The model used pressure transducers along the pipeline, that were connected to digital-to-analog and analog-to-digital converters. Through reading and analyzing transient pressure plots, the model can detect leaks with the size of 5% up to 50% of the flow rate. Furthermore, the model can identify the leak location using the pressure wave velocity. The identified leaks were within 5 meters of the exact location of the leak. The accuracy of pressure transducers can be enhanced by coupling it with the Frequency Response Method. The proposed model was capable of successfully detecting and locating small leaks (i.e. leaks as small as 0.5% of the average flow rate). This method relies on measuring pressure values and fluctuations as well as flow rate fluctuations. The collected data was analyzed by the frequency response method and used equations of continuity and momentum. The model presented equations that were capable of identifying the location of a leak for one or more anticipated leaks. This method cannot be applied to materials that are beyond a specific range of material friction. Furthermore, 0.5% of flow rate is a tremendous value of lost water in large transmission pipelines (Mpesha et al. 2001).

As pressure is a known measure, improvements are commonly needed in the analysis stage. For example, by analyzing the pressure data provided by the transducers using Hilbert transform and

Hilbert-Huang transform, the approach can detect leaks within a pipeline segment by recognizing the reflection in the analyzed signal data. However, this technique cannot pinpoint the exact leak location (Ghazali et al. 2011). To adopt this technique to modern day needs, real-time capabilities were added by Sala and Kolakowski (2014). The model relies on the transmission of pressure measurement throughout a network that uses a GSM platform. Then, the collected data is analyzed by a software and converted into pressure data, to predict the leak location, using a set of developed equations.

Ground Penetrating Radar

Stampolidis et al. (2003) have worked on a case study of Ground Penetrating Radars (GPR), proving that they can detect underground water pipelines under ideal GPR working conditions. The case study includes a large number of field experiments and a significant amount of collected data. A leak was identified at the existence of any disruptions in the coloring pattern, usually red, within the images derived from the GPR readings. The findings were compared to the exact leak location and the radars were assessed to have accurate leak detection under ideal weather and operational conditions. Improving the readings of GPRs can be conducted through radargram refinement, which converts the raw images collected by the GPR into usable data that are more effective in leak detection and localization (Al Hawari et al. 2016). GPR readings can be further improved when coupled with infrared readings. In this case, the GPR images can be used as a second layer of images to overlay the infrared images, thus providing a more accurate leak location estimate. This technique provides a small margin of error ranging between 2.9% and 5.6% (Atef et al. 2016).

Vibro-Acoustic Methods

Vibro-acoustic methods are arguably the most famous leak detection and localization technique currently. Those methods rely on the collective power of vibration detection technologies and acoustic listening technologies. The term ‘vibro-acoustic’ comes from using the two technologies together. Often, acoustic listeners would detect a specific leak location and the data derived from vibration measurement devices helps identify the accuracy of that value. Therefore, various researchers were directed at improving the vibro-acoustic technologies and empowering them. Vibro-acoustic technologies can be cost-efficient and simple for multiple types of pipes, including plastic pipelines. In 2004, Hunaidi et al. developed a mobile leak detection system, named LeakFinderRT. The model used a computer software directly connected to acoustic listeners coupled with wireless transmitters and low-frequency vibration sensors. The software receives the noise signal from the transmitters and processes it. In the next step, an enhanced cross-correlation method is used to detect the leak and localize it. However, the system needs further testing in the case of large pipelines. To operate, the system requires three main inputs: pipe type, pipe diameter, and sensor-to-sensor spacing. The author displayed that vibration sensors -accelerometers- can be very effective in detecting leaks in plastic pipelines. Additionally, the model relies on leak suspicion for the user to go to a suspected leak site and detect a leak. Even then, non-linear pipeline shapes may cause the system to produce incorrect results (Hunaidi et al. 2004). In the previous model, the cross-correlation method is used to detect the behavior of water pressure, flow velocity and pipeline vibration (Gao et al. 2005). The model showed that acoustic detectors are efficient when there is minimal noise, yet correlation can be vastly improved when vibration signals, derived from accelerometers, were utilized to

complement the correlation technique. This model also included a set of experimental field testing to validate the theoretical results.

Several researchers referred to the vibro-acoustic approach for leak detection as a highly accurate technique for estimating and pinpointing a leak. For instance, Brennan et al. (2007) present a model using a vibro-acoustic approach. The authors use the time domain and the frequency domain to pinpoint leaks within pipeline systems. Acoustic listeners and accelerometers are used to detect the signals of leaks. The received signals were filtered to eliminate noise and they are processed by a privately developed software to be converted to frequency domain figures and time domain figures. Then, the time delay is estimated using the time between the signals, to assess the speed of the signal and triangulate the leak location. Real life experiments were performed using fire hydrants as access points and, as the results show, the difference in delay between the time domain and the frequency domain is negligible. Vibro-acoustic detection was improved via collecting signals in the time and frequency domains and then filtering the noise out of the collected data via wavelet transform. This modification helped improve acoustic technologies in leak detection and raise their accuracy through decreasing the possibility of having false alarms (Meng et al. 2012). For more accurate estimates of leaks, Khuleif et al. (2012) proposed the utilization of in-pipe hydrophones as complementary measurements for acoustic detectors or even as a standalone leak detection system. The authors developed a software to process hydrophone data collected from the experiments. The research also concluded that the leak signal is more prevalent with the flow rate rather than against the flow of water.

As real-time monitoring is gaining more popularity, real-time monitoring enhancements can be coupled with vibro-acoustic technologies to provide over-the-clock monitoring of water

networks to detect sudden failures, blockages and third-party interferences such as sabotage. To accomplish the model, flow-induced vibrations were collected using accelerometer sensors and then they are analyzed. The relationship between the pipe over-all-status and the flow vibration signals acquired by the accelerometers is studied by a set of lab experiments. The researcher concluded that there is a strong relationship between the vibration signal of the flow and the flow rate. The research also revealed certain limitations due to the interference of external factors that can distort the signal (Awawdeh et al. 2006). Real-time models are developed using new iterations and device under the vibro-acoustic category such as Stoianov et al. (2007), who proposed a system to help in near real-time monitoring of water networks as well as water network control; the system was dubbed “PIPENET.” The system is composed of a field-deployable combination of hardware and software capable of measuring and collecting hydraulic, water quality, acoustic and vibration data. The data was collected over twenty-two months of trial in the city of Boston in the US. The collected data was fed into cross-correlation algorithms and signal analysis software to detect and locate leaks in the network under study. Besides, lab experiments further validated the model capabilities. The system relied mainly on the analysis of pressure signals, vibration signals and acoustic signals. The system showed promising results; yet, according to the authors, more work is required to develop more accurate results and minimize the number of false alarms given by the system. Relevant research suggest that vibration monitoring and acoustic signal monitoring can be separated, and each technology is capable of detecting and locating leaks on its own. Mostafapour and Davoudi (2013) studied the impact of leaks in pressurized pipelines using the stress waves from the energy loss caused by leaks. The model uses acoustic methods to model the acoustic emission that is created by vibration in the pipeline because of leakage. For more accurate results than in conventional

approaches, this model used Fast Fourier Transform. The mean level of error between theory and the experiments was averaged at 6%. One of the models that rely completely on wireless acoustic technology came in the form of a noise logger leak detection and localization model developed by El-Abbasy et al. (2016). The authors propose that noise loggers can go beyond leak detection and are capable of leak pinpointing. Multiple experiments were conducted and data was collected from the acoustic listeners. Regression analysis and Artificial Neural Networks (ANN) was used to analyze the data. The results show that both models presented high levels of accuracy, with accuracies of 88% and 93% from the original leak location for the regression analysis and ANN models respectively.

Accelerometer-based approaches have received more attention in recent years, as a complete leak detection system, using their capabilities to detect vibration signals that are emitted by leaks. Shinozuka et al. (2010) suggested using accelerometers to measure the variation in acceleration forces (g-force) and record those variations to identify damaged areas and map the network. Regarding standalone vibrational models, Almeida et al. (2014) developed a technique for the interpretation and assessment of the data coming from accelerometers, which monitor water pipelines by using a set of mathematical models. The model showed enhanced accuracy in identifying the frequency range of accelerometers, identifying time delays between the two sensors and estimating the wave speed of propagation. The correlation formula can use the derived values to pinpoint the leak. Figure II-6 shows the experiment by Almeida et al., where one sensor is located on the right side of the leak and the other on the left side of the leak. The sensors are placed on the valves connected to a buried plastic pipeline. A leak was simulated at a random distance in the pipeline between the two sensors in order to study the variation in vibration signals and accordingly create a leak detection and pinpointing model.

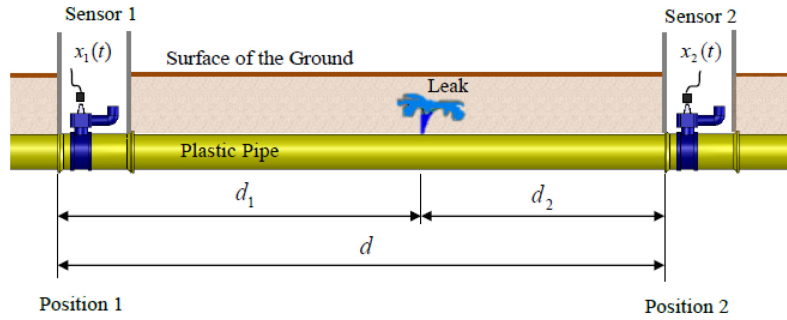


Figure II-6: Schematic of buried water pipe in which sensors are positioned at access points. (Almeida 2013)

The correlation formula, in this case, is displayed in Equation (II-1). The equation relates d_2 , that is displayed in the figure, to the total distance d by using the time difference between the two sensors and the normal wave speed in the pipeline. The wave speed in the pipeline is expected to vary from one material type to another and it is usually assumed to be 400 m/s in ductile iron pipelines. After determining d_2 , d_1 can be determined via subtracting the value of d_2 from the total distance d .

$$d_2 = \frac{d - ct_0}{2} \quad (\text{II-1})$$

Where:

c : wave speed

d : the total distance between measurement positions

d_2 : distance from sensor 2 to the leak

t_0 : the time delay between sensor 1 and sensor 2

Additionally, when it comes to vibration analysis, Martini et al. (2015) presents a model that relies on vibration signals in PVC pipelines. The model presented a vibration signal analysis approach to decipher the signals and convert their values to a comprehensible index. The model

proved to be accurate in identifying the slightest leaks in PVC pipelines. This model was improved by El Zahab et al. (2016) through further experiments. The model used the index already developed by Martini et al. (2015) to develop a regression model that can pinpoint the leak in PVC and ductile iron pipelines with an average accuracy of 95%.

Fiber Optic Sensors

Inaudi et al. (2007) has made a step forward in the volume balance technique that can detect small leaks in water transmission systems. The system relies on fiber optic sensors throughout the network. The sensing systems are based on Brillouin and Raman scattering. The system can detect leaks in fluid, gas and multiphase transmission lines. Using this approach, the system can locate a leak with an accuracy of one to two meters by measuring variation in distributed temperature and strain. When it comes to limitations, the system allows room for false alarms as both heat and strain can be affected by external factors. Besides, the system installation might not be affordable for small to medium municipalities.

Infrared Thermography

Fahmy and Moselhi (2009) presented a model that integrates the imagery collected from an infrared surveyor with computer-based computational models to monitor underground pipelines and locate any possible leaks. The research presents experimental work over the span of two years, utilizing the infrared technology. Based on the fieldwork, a localization equation was presented in the model to locate possible leaks. Upon comparing the model to acoustic and vibratory models, the model has proved viable and accurate. The limitations of the infrared technology itself limit the model in terms of working temperatures and certain times of the day

avored over others. Infrared detection can be further improved when combined with statistical analysis and prediction techniques. Statistical analysis can help define leak-related patterns in images and leak-related temperature thresholds. Under ideal operating conditions the coupling of IR thermography with statistical analysis can provide leak detection and localization models with 95% accuracy (Al Hawari et al. 2015).

Other Technologies

In this section, various models from different backgrounds are investigated and the approaches and developmental technologies are explored. Those models might not be correlated entirely but they revolve around leak detection models that cover the phases of localizing, locating and pinpointing. Colombo and Karney (2002) have developed a model that relates consumed energy to the energy expenditure and costs of pumping water through water pipelines. The model presents a set of equations that use the changes in energy consumption of pumps to identify the existence of the leak. Then, the leak can be located using the energy consumption envelope. The authors highlight the importance of the immediate leak repair and a prioritization scheme for the leak repair order. The model also confirms that leaks create costly impacts on the level of water resource and the age of the water network.

Whittle et al. (2010) presented another wireless monitoring system for water mains by the name “WaterWise@SG”. The research objectives here were to (1) develop a system to provide high-quality data for low costs in a remote manner and (2) monitor and assess the parameters that govern the water quality and hydraulic quality. The devices in this research consisted of pressure sensors, hydrophones and flow meters. Besides, the devices had 3G and Wi-Fi interfaces to send the acquired data. The data was received via an automated platform and was analyzed by a

software privately developed to model the state of the water network. When it comes to its limitations, the device itself is too big for small networks. Moreover, the algorithms and the software need to be more accurate and efficient.

Li et al. (2011) have used a geographic information system (GIS) platform along with the district metering area technique to create an integrated system for detecting, early warning and leak control in water mains. The system was tested in Beijing on an experimental scale and was capable of detecting 102 non-obvious leaks. Additionally, a leak prediction algorithm was developed using historical data to pinpoint the next possible leak. However, this system still requires a lot of testing and experimenting to increase the accuracy of the system. Another wireless network for monitoring underground transmission assets was presented by Sun et al. (2011) namely “MISE-PIPE.” The presented system relies on a wireless sensor network that is based on magnetic induction wireless sensors. The model presented a low-cost solution with real-time leakage detection and localization capabilities using a magnetic induction waveguide technique. MISE-PIPE consists of two sets of pipelines, a set inside the pipeline and a set on the surface of the pipeline. The set consists of magnetic induction sensors, acoustic sensors, pressure sensors and soil property sensors. This novel technique is promising, but it requires further research on real-world applications and sensor placement reconfiguration in the pipeline.

II.1.5 Accelerometer Signal Analysis

Throughout this research endeavor, twenty-three attempts for the analysis of leak signals through accelerometers were conducted. The approaches included Fourier Transform, moving average calculations, and basic value measurements. The approaches did not provide a significant assessment for the signal that can be used in model development. Additionally, the models did

not highlight the change in condition effectively. In 2015, a research paper was published by Martini et al. that addresses the issues tackled in this research work. Martini et al. (2015) proposed a model to use accelerometers for leak detection. They also suggested an approach for analyzing the signal received from the sensors. Their mathematical analysis approach can be summed in four steps as follows:

1. Determining acceleration reading per second in (g).
2. Each $t = 100$ second, the readings are collected and their standard deviation is determined.
3. After several hours of monitoring, the lowest ten standard deviations are averaged to determine the lowest monitoring index as shown in Equation (II-2).

$$MI_j = \text{mean}(\sigma_j, 10) \quad (\text{II-2})$$

4. A value named Monitoring Index Efficiency (MIE) is determined by dividing the current monitoring index of any instant with the lowest monitoring index in no leak state of a given duration $MI_{j,\text{no leak}}$ as illustrated in Equation (II-3). This equation allows the establishment of sensor-specific values. For example, MI_0 can be unique for each sensor and the readings, hence taking into consideration any pre-existing conditions and external factors. Martini initially utilized the maximum monitoring index of a duration, whereas MIE is determined each $t = 100$ second in this research.

$$MIE_x = \frac{MI_i}{MI_0} \quad (\text{II-3})$$

Where:

- MIE is the Monitoring Index Efficiency.
- x can be either L or R, representing left or right sensor respectively.

- MI_0 is the lowest recorded monitoring index at no leak state.
- MI_i is the monitoring index of a given signal at time i .

II.2 Prioritization of Independent Events

This thesis also tries to tackle the problem of leak repair prioritization and resource allocation for repair projects. Thus, here, leak repair projects are considered independent events, i.e. not related in terms of succession. Moreover, all leaks are assumed to be deteriorating at an individual rate determined by each case and through historical data. Accordingly, in this section, previous work on prioritization approaches of independent events under constraints are discussed. Some researchers attempted to distinguish independent event prioritization approaches from regular scheduling approaching and dubbed them the name “Priority Algorithms.” Davis (2003) identified priority algorithms and their role in solving prioritization problems. The report identified multiple algorithms capable of solving prioritization problems, mainly greedy algorithms, genetic algorithms, adaptive priority algorithms and dynamic programming. The author also distinguished between fixed priority algorithms and adaptive priority algorithms. An early approach for optimizing independent event ordering was presented by Coloroni et al. (1992). Their proposed approach used evolutionary Genetic Algorithms (GAs) to solve the timetable problem. This problem is limited by school hour constraints and teacher schedule constraints and the objective was to remove conflicts in classrooms and to remove the course overlaps. The authors simulated the problem as a multi-constrained, NP-hard, combinatorial optimization problem. Using genetic algorithm with local search as well as genetic algorithms without local search, the authors have concluded that genetic algorithm presents more flexibility in defining constraints and outperforms handmade timetables and simulated annealing

timetables. Scheduling problems can be solved using a resource-driven model that is capable of optimizing the schedules of linear projects by utilizing a dynamic programming formulation and heuristic rules, as proposed by Moselhi and Hassanein (2003). The model accommodates repetitive and non-repetitive optimization activities to develop practical and near-optimal schedules. A software was developed to embody the theoretical concepts of the model and develop case studies. The model has proved to be capable of optimizing construction time, cost or both under a cost-plus-time bidding environment.

Prioritization models are required not only to prioritize but also to optimize expenditure and cost. For example, Morcous and Lounis (2005) have presented a model that is capable of minimizing life-cycle cost of infrastructure through predicting deterioration in infrastructure networks using genetic algorithms and Markov chain networks. This model is important in this research as it shows previous work predicting and assessing multiple maintenance alternatives for a specific network within a specific timeline. To prioritize independent events, a scale of measurement is required. Elbehairy et al. (2006) has introduced a repair prioritization approach for bridge deck repairs. Two evolutionary algorithms are used in this approach to optimize the order within the repair order. These two algorithms are genetic algorithms and shuffled frog leaping algorithms. Each repair event is identified by three aspects as follows: (1) Expected deterioration, (2) Cost and (3) Repair impact. This research has concluded that both optimization approaches are equally suitable to optimize bridge repair prioritization. Another prioritization model, this time for pipe replacement prioritization, was presented by Giustolisi and Berardi (2009). The authors identify three main aspects of leak repair: (1) Economic aspect, (2) Reliability aspect and (3) Water quality aspect. The main aim of the model is to develop a plan that maximizes the effectiveness of each pipe rehabilitation to help the decision maker identify how to tackle the

replacement plans within an infrastructure network. The primary tool for the development of this model was a multi-objective evolutionary optimization with the best results obtained when the multi-objective genetic algorithm is utilized. The research above proves the need for an approach that can help the decision maker plan for the long-term changes within multiple similar projects. A dedicated approach to the prioritization of events, leaks specifically and assess the best manner to allocate resources is not directly available.

II.3 Model Development Techniques

In this section, various model development techniques are explored for their uses and algorithms. The four techniques to be presented are: (1) Artificial Neural Networks (ANN), (2) Genetic Algorithms, (3) Regression Analysis and (4) Support Vector Machines (SVMs).

II.3.1 Artificial Neural Networks (ANN)

The biological term “neural network” refers to an interconnected grouping of simple processing units, i.e. neurons, that are coupled via axons and dendrites. The term “artificial” refers to an algorithm that simulates a network of simple processing units and has an architecture similar to that of the naturally occurring neural networks. In an artificial neural network, synapses are processing units that carry a pure, usually binary, value, referred to as a weight. Each synapse operates via a threshold logic unit that reacts based on the incoming input values provided by the input dendrite or link. If the input is sufficient enough to tilt the threshold, then the processing unit will output a high value, which is usually one; otherwise, the output would be zero. From a machine learning perspective, artificial neural network is a supervised learning technique. A

supervised learning technique requires a set of learning data to which the algorithm tries to adapt its neuron layers. The developed model can be validated against the testing data (Gurney 1997). Artificial neural networks have recently improved, and interest has grown in feed-forward neural networks, also referred to as “Perceptrons.” Figure II-7 shows that the structure of a neural network algorithm is composed of three essential layers. The first layer is the input layer where all the required learning input is provided to the algorithm. The second layer is the hidden layer, where several layers can be established to fit the needs of the user. Inside the hidden layer, the processing stage occurs using a learning algorithm such as backpropagation. The final layer is the output layer where the results of the model are presented and the quality of the solution is assessed (Müller et al. 2012). Artificial neural networks are widely applied in the leak detection field due to the algorithm’s capability to identify existing patterns that are not readily recognizable. Furthermore, artificial neural networks are capable of unearthing relationships between factors and presenting accurate solutions. Belsito et al. (1998) utilized artificial neural networks to detect leaks in liquefied gas pipelines. The inputs for the model are pressure, flow and energy measurements. To recognize a leak pattern, several experiments were done and simulated leaks as well as a no-leak pipeline, i.e. an ideal state pipeline were used. The developed model can detect leaks as small as 1% of the total flow rate within a time frame of 100 seconds. However, the success probability of the model is around 50%, which is considerably low.

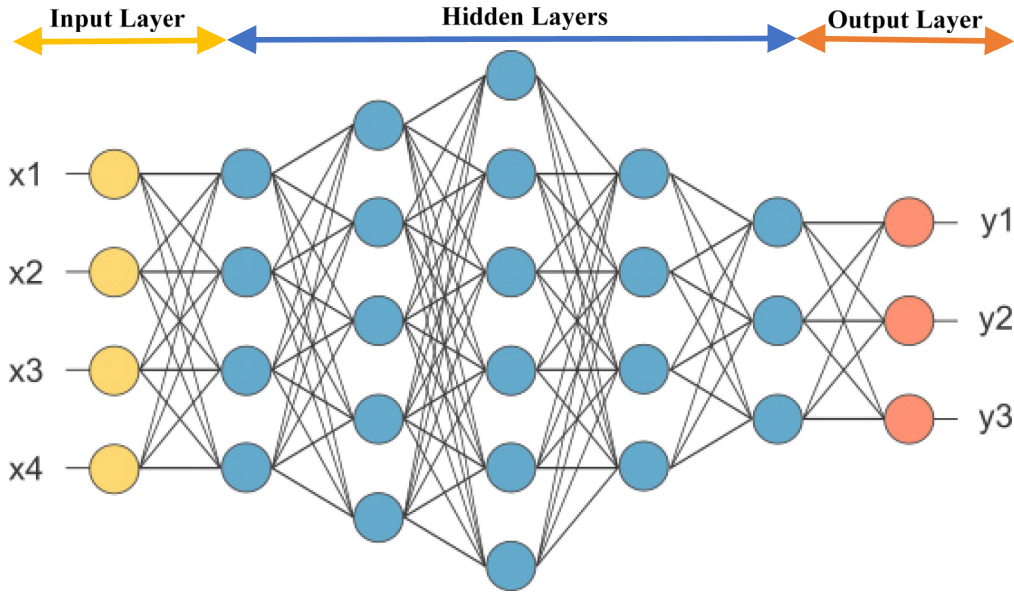


Figure II-7: Artificial Neural Network Structure

With technological advancements in sensors and artificial neural networks, the above-mentioned model was improved by Mounce and Machell (2006) who used flow and pressure measurements besides two neural networks architectures, static and time delay to identify and classify time series patterns related to leaks. Based on the developed classifications, the existence of leaks can be determined if any time series fits the pre-determined patterns. Moreover, through a field trial and a set of experiments, the authors have concluded that artificial neural networks rely on sufficient learning data as well as a high data quality to be effective. They have also identified the time delay neural network as the better counterpart of the static neural network regarding result quality. Besides, several researchers tried to identify the most suitable ANN function for leak detection systems. For example, Santos et al. (2013) have compared two artificial neural network functions, multilayer feedforward and radial basis function network. They did experiments and leak simulation on a 60-meter galvanized iron pipeline and used acoustic listeners or microphones. The collected data was fed to the two algorithms and the results were compared. The multilayer feedforward network, also referred to as multilayer perceptron

network, performed slightly better than the radial basis function in leak detection using acoustic signals within a 60-meter galvanized pipeline.

The previous attempts at coupling acoustic data with ANN algorithms highlight the potential of having an automated leak detection system based on acoustic listeners supported by ANN algorithms. One of the recent approaches in this domain is proposed by El-Abbasy et al. (2016). They show the application of noise loggers on 6-meter ductile iron pipelines of one-inch diameter size. The experiments are fed to both a regression platform and a neural network platform and the results are compared. The ANN model outperforms the regression model and presents a model capable of providing a solution with a 93% validity as well as a mean square error close to zero. Based on what was discussed so far in this section, the artificial neural network algorithm can provide accurate and robust models to help in the process of leak detection mainly in ductile iron and galvanized iron mediums. However, the ANN algorithm is seldom capable of providing equations to be immediately applied because the ANN algorithm performs as a black box, with the layers in the algorithm remaining unknown.

II.3.2 Genetic Algorithms

Genetic Algorithms (GAs) were invented by John Holland in the 1960s and are currently used to solve a variety of optimization and evolution problems. GAs rely on mimicking the concepts of natural biological evolution and utilizing the power of biological adaptation. Application of GAs and other concepts of evolutionary computation have grown in the engineering field (Mitchell 1998). Inspired by natural evolution and the graphical summary provided in Figure II-8, the algorithm starts with a set of possible individual solutions. This set is referred to as a population of solutions. Each solution is then assessed for its level of fitness using a fitness function.

Through the level of fitness, original solutions become parent solutions and can reproduce and create a new generation of solutions. The new generation has the same size as the old generation and each new solution has mixed properties from its respective parent solutions. Eventually, if the problem is well designed, the algorithm will converge to an optimal solution (Goldberg 2006).

GAs are frequently used in the field of prioritization and schedule optimization. Researchers have presented multiple possibilities with evolutionary algorithms capable of providing solutions for complicated multi-objective optimization problems, e.g. scheduling, and event prioritization problems. Cai and Li (2000) proposed genetic algorithms to solve a scheduling problem. They presented a multi-objective genetic algorithm to optimize the schedule of 624 staff members based on three predefined criteria by the user. The algorithm proved effective in organizing shifts, optimizing the schedule and reducing the costs through decreasing the required number of staff members to 593 individuals. The research works reviewed in this section and those presented in section II.2 show that evolutionary algorithms in general and specifically genetic algorithms offer a feasible and practical solution to prioritization problems and are capable of providing adequate and reliable optimal schedules. On the other hand, evolutionary algorithms can be time-consuming as they are not explicitly designed to solve prioritization problems.

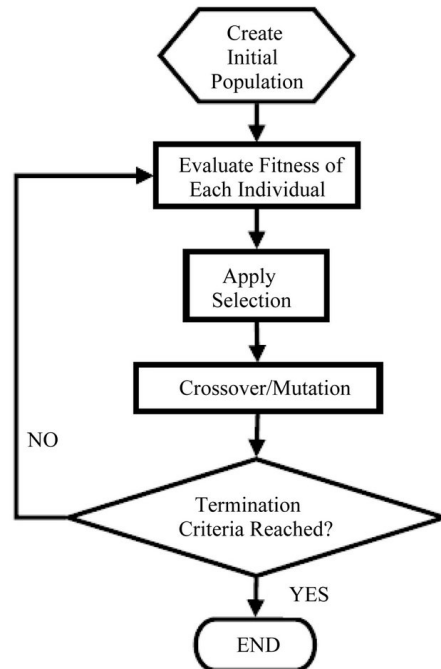


Figure II-8: Generalized Genetic Algorithm Flowchart

II.3.3 Regression Analysis

Regression analysis is a statistical modeling technique. The primary goal of regression analysis is to develop mathematical models that are capable of correlating and defining, in a sensible manner, the behavior of selected variables (Sykes 2007). Both the quality of the developed regression models and their applicability to reality are directly related to the quality of the provided data. Two main types of models can be developed via regression analysis. The first type of models, named linear models, are formed by simple input-output correspondence and the development model would be a linear equation. The second form of models, named non-linear models, are deemed more applicable to real-world problems as they explore multiple possible relationships beyond linear solutions between the input and the desired output (Rawlings et al. 1998). Regression analysis aims to develop a line of best fit between the inputs and the outputs. Therefore, the analysis presents an equation of the developed model using the provided training

data and the quality of the model can be assessed through indices such as mean square error, r-squared, p-value, mallows C_p and s-value (Sykes 2007).

In terms of leak detection, regression analysis is widely utilized due to its powerful pattern recognition capabilities and its capability to provide an equation for predicting leak locations. Wang et al. (2002) suggest regression analysis to analyze and assess transient flow to detect leaks. The linearized assessments and experiments can determine the governing dampening equations, which reduce the signal emitted by the leak, and thus locate leaks as small as 0.1% of the total flow rate. Another advancement in regression analysis for leak detection is presented by El-Abbasy et al. (2016). They correlate the percentage of the leak location from each acoustic sensor to the level of noise heard by the left and right sensors. The presented approach is able to develop models that can pinpoint the leak with an accuracy of 88%, with a mean square error that is close to zero. However, when the regression model was compared with the ANN, the values of the ANN had a higher accuracy by 5%. Regression analysis is a technique of correlating relationships and identifying patterns and eventually developing models that represent the relationships under study. The mathematical models presented by a well-developed regression model can give better understanding of incomprehensible relationships.

II.3.4 Support Vector Machines (SVM)

The Support Vectors Machine (SVM) algorithm is a supervised learning classification algorithm first proposed by Cortes and Vapnik (1995). The algorithm was developed to serve as an intelligent classification technique. One of the main capabilities of SVM is to create thresholds, that can aid in decision making, and solutions for binary classification problems. The SVM algorithm can be divided into two main categories: (1) Linear SVM for linearly separable

problems and (2) non-linear SVM for problems that are not linearly separable and require much more complex solutions (Fletcher 2009). The SVM algorithm has also been utilized to identify leak locations as proposed by Mashford et al. (2009). The authors have used pressure data and flow rate data to detect and pinpoint leaks using the SVM algorithm as a classifier and as a regressor. The data was collected through multiple experiments, via a hydraulic modeling system that is dubbed “EPANET.” The developed regression model via SVM algorithm had a high r -squared that is close to 100% and a mean square error close to zero. The model relies heavily on the quality of the provided data and the sensitivity of the used sensors.

Besides, the SVM algorithm is functional in leak detection when coupled with other concepts such as the rough set theory as presented by Mandal et al. (2012). The rough set theory was used to create a set of rules that would exclude any predetermined and easily identifiable non-leak states or leak states to guarantee that the SVM algorithm would not be confused by any anomalies and then, a trained SVM algorithm was used to classify the cases that were not ruled out by rough set theory. This approach has multiple advantages including high accuracy and capability to replace mass balance and pressure point analysis. It also provides a detailed simulation of the pipeline being studied and can be effective to be utilized in old pipelines. The SVM algorithm is deemed to be a robust supervised classification and regression tool that is capable of differentiating leak states and identifying leaks’ location, given that the data is well constructed.

II.4 Limitations of Current Approaches

In this section, the limitations of the currently used approaches are discussed in terms of technologies and prioritization approaches. Leak detection technologies are compared and assessed against one another.

II.4.1 Limitations of Current Technologies

From the perspective of this research, a specific guideline is necessary to define leak detection systems and to identify what is expected from such systems. Geiger (2006) identified four main aspects to evaluate the performance of leak detection systems. The four main aspects are (1) reliability, which represents the extent to which the system is capable of detecting a leak and the accuracy of the received information, (2) sensitivity, which shows the capability of the system to identify leaks of various sizes ranging from big leaks and bursts to small leaks, (3) accuracy, which, unlike reliability, provides extra information such as leak location and leak size and (4) robustness, which displays the ability of the system to provide reliable data even under changing conditions within the network or in the case of data loss. Additionally, Geiger identified that externally-based leak detection sensors can pinpoint leak accurately, whereas, with internally-based leak detection systems, leak locations are estimated and are rarely as accurate as the values derived from externally-based leak detection sensors. When it comes to the costs, externally-based leak detection systems are considerably expensive and complicated to install. Thus, they are not used as much as internally-based systems that cost much less. Externally-based leak detection systems, as stated in section II.1, rely mainly on sensors for collecting various types of signals and data. However, externally-based leak detection systems have proven to be the most capable of leak pinpointing. Therefore, the limitations of such systems can come from the

technologies they utilize. Thus, in this section, a brief discussion and a comparison of technologies and their limitations will be discussed.

Lately, multiple technologies are utilized to minimize the damage created by the leaks. One of the most popular technologies is acoustic detectors or listening devices. Listening devices rely on identifying the sound emitted by the leaks to detect a leak and, in some recently developed models, pinpoint the leaks. The primary tools in this category are geophones and hydrophones, described in section A.1 of the appendix. Those devices listen to audible sounds in the network mainly using listening sticks or microphones. Besides being popular, acoustic technology encounters multiple limitations that impair their effectiveness as follows: (1) Acoustic devices must go over or in the pipe, which might require opening the pipeline or drilling and digging sometimes; (2) acoustic devices are affected by ground dampening on the signal and thus their accuracy is reduced, (3) acoustic devices rely on well-trained operators to detect and pinpoint the leaks while listening and the human operators must be able to distinguish the signal from the background noise picked up by the devices pickup; (4) acoustic devices need to be close to the leak to be able to detect it, not functioning efficiently in plastic pipes such as PVC (Echologics Inc. 2006; United States Environmental Protection Agency 2009).

Another popular technology, considered an improvement in acoustic phones, is acoustic devices with correlation techniques such as noise loggers. Such devices rely on placing two acoustic listeners on opposite sides of the expected leaks with their data input into a correlator, as described in section A.2 of the appendix. This technology has shown promising results and received multiple contributions. Nevertheless, they are costly to utilize and they require the anticipation of the leak location to be confirmed. Besides, acoustic devices with correlators face difficulties in correlating small and low-sound leaks and tend to struggle detecting leaks within

PVC pipelines and large diameter pipelines (Echologics Inc. 2006; Datamatic Inc. 2008; El-Abbasy et al. 2016).

Infrared thermography has been widely used for leak detection through sensing the infrared radiation and temperature change within surfaces where leaks are expected. Section A.3 of the appendix provides more details on this technology. However, infrared devices can be expensive and require well-trained operators for good results. Moreover, the accuracy of infrared devices can be impaired by external factors, mainly weather conditions that limit the operational capabilities of the technology to a specific temperature range (Echologics Inc. 2006; Fahmy and Moselhi 2009; Varone and Varsalona 2012). Another form of leak detection technologies is chemical leak detection. Chemical leak detection is performed using tracer gasses injected into the network and escaping through leaks to be detected at the surface as described in section A.4 of the appendix. Chemical leak detection, however, faces multiple challenges as it takes a substantial amount of time to detect a leak and thus a large number of resources and huge funds are needed. Moreover, the accuracy of chemical leak detection is affected by factors such as depth and inner soil paths that might render the leak to go out of another location or dissipate through the soil (Echologics Inc. 2006; KVS 2015).

Another technology widely used in leak detection is Ground Penetrating Radar (GPR). This device relies on emitting radar waves into the soil and generating an image of the soil based on the reflection of the emitted signals, as illustrated in section A.5 of the appendix. However, the images derived from the GPR data are hard to interpret and require extensive research to become viable (Eyuboglu et al. 2003; Echologics Inc. 2006; United States Environmental Protection Agency 2009).

Multiple devices have recently been developed to solve the leak detection problem. This class of devices is known as leak detection robots that include devices such as Smart Ball, Beaver, Explorer and Pearpoint. As illustrated in section A.6 of the appendix, most of these devices are able to venture through pipelines on their own with little or no human aid. However, these devices are, in some cases, exclusive to the developer company. Additionally, leak detection robots fall under the risk of getting stuck in the pipeline and having a moderate accuracy level compared to other available technologies (Schempf et al. 2003; WCT Products 2015; Puretech Ltd. 2015).

Micro-electro-mechanical sensors (MEMS) have been extensively used in recent years as real-time leak detection devices. MEMS belong to a broad category of sensors, which contains various types of sensors as described in section A.7 of the appendix. MEMS in this research are accelerometers, as devices that can read vibration signals. Accelerometers still require more research in terms of plastic pipelines and metallic pipelines to become a standalone technology for leak detection., Studies are also required in terms of the signal interpretation to eliminate any noise within the signal, mainly due to the high sensitivity of accelerometers (Kim et al. 2011; MEMS and Nanotechnology Exchange 2015; El-Zahab et al. 2016).

Table II-2 provides a comparative analytical summary of all the aforementioned leak detection technologies and assesses the technologies based on their cost, accuracy, capabilities of detection in metallic and non-metallic pipelines, market availability, traffic interruption and susceptibility to external factors. Another factor is permanent installation determining whether a system is permanently installed in the network to provide real-time data or it is brought in when a leak is detected. For example, micro-electromechanical sensors (MEMS) provide a low-cost solution with high accuracy and they can be installed permanently in the network. However, MEMS are

still experimental when it comes to experimental pipelines and plastic pipelines and their functionality needs further research. Leak detection using MEMS is affected by external factors that need to be addressed in signal analysis, yet this detection does not interrupt the traffic flow. Besides, the table indicates that MEMS are readily available in the market for purchase.

Table II-2: Comparison of the Available Technologies against Multiple Criteria.

Technology	Cost	Accuracy	Perman ent	Metallic Pipes	Non- Metallic Pipes	Market Available	Interru pts Traffic	External Factors
Geophone/ Hydrophone	Low	Low	No	Yes	Yes	Yes	Yes	Yes
GPR	High	Operator dependent	No	No	Yes	Yes	Yes	Yes
Infrared	n/a	Low	No	Yes	Yes	Yes	Yes	Yes
Leak noise Loggers	High	High	Yes	Yes	No	Yes	No	Yes
Smart Ball	Low	Moderate	No	Yes	Yes	Company Only	No	No
MEMS	Low	High	Yes	No	No	Yes	No	Yes

II.4.2 Limitations of Prioritization Approaches

In the field of single event prioritization, multiple contributions have been made (as mentioned in section II.2.), using evolutionary learning algorithms and more specifically genetic algorithms. Although beneficial and powerful, genetic algorithms have drawbacks in terms of computational efficiency and time efficiency. Evolutionary algorithms are considered time-consuming and might deflect away from the optimum solution to near-optimal solutions (Davis 2003). Additionally, evolutionary algorithms can be inefficient in handling large quantities of events without affecting their computational performance and lowering their respective speed (Colorni

et al. 1992). Finally, most evolutionary algorithms tend to utilize static schedules to create optimally prioritized schedules. However, in the case of prioritizing leak repairs, there is a need for more dynamic approaches to consider the worsening of the condition of a leak that is to be repaired. Therefore, an algorithm with minimal simulation and prediction capabilities to prioritize events as they move through time is highly needed (Colombo and Karney 2002).

CHAPTER III: ACCELEROMETER BASED MODELS

III.1 Accelerometer Real-time Monitoring Model

The second objective of this thesis is to develop and validate technology-based leak detection and pinpointing models. In this chapter the review constructed in Chapter 2 is built upon to select accelerometers as one of the main devices for development. Therefore, Chapter 3 goes over the selection and development of the leak detection models that utilize accelerometers (MEMS). The review of available leak detection technologies is followed by a set of criteria established to determine the technology that are developed in this research. As illustrated in Figure III-1, these three primary criteria are:

- *Real-time Capabilities*, highlighting the technology's ability to send almost instant monitoring data to control centers.
- *Cost-Effectiveness*, representing the expected costs per network for a specified technology based on the collected data.
- *Room for contribution*, identifying the amount of research still required for the specified technology.

Based on those three criteria and the literature review, microelectromechanical sensors (MEMS) and accelerometers are identified to have high real-time capabilities and an affordable cost and they still require a significant amount of research. Therefore, accelerometers are selected to be the subject of this research.

The selection of the device is the first stage in the process of developing the real-time monitoring system model, as shown in Figure III-2. After accelerometers are selected, the signals they emit need to be deciphered. Then, the signals should be processed into meaningful values and indexes to help identify and pinpoint leaks. Following analyzing the signals and establishing leak

identifying values, the data is used to develop a model capable of informing the user of the existence of a leak. To improve the model, additional experiments have been done to determine if the model can detect the size of the detected leak. This way, the accelerometers are able to pinpoint the origin of the signal (i.e., the leak location). Knowing the possible areas of leakage and their respective monitoring index values, the values are entered into a machine learning model to automate the process and determine the leak location. The final step is to identify how the signal moves in the leak by utilizing the discovered leak locations and the values sensed at far away sensors. These models will be further illustrated and explained in the following sections. Additionally, at this stage, the methodology used for acoustic loggers is identical to that of accelerometers except when it comes to signal analysis, where sound waves undergo Fourier transform and signal amplitude analysis. This is mainly due to acoustic loggers recently added to this research project, that are still under development.

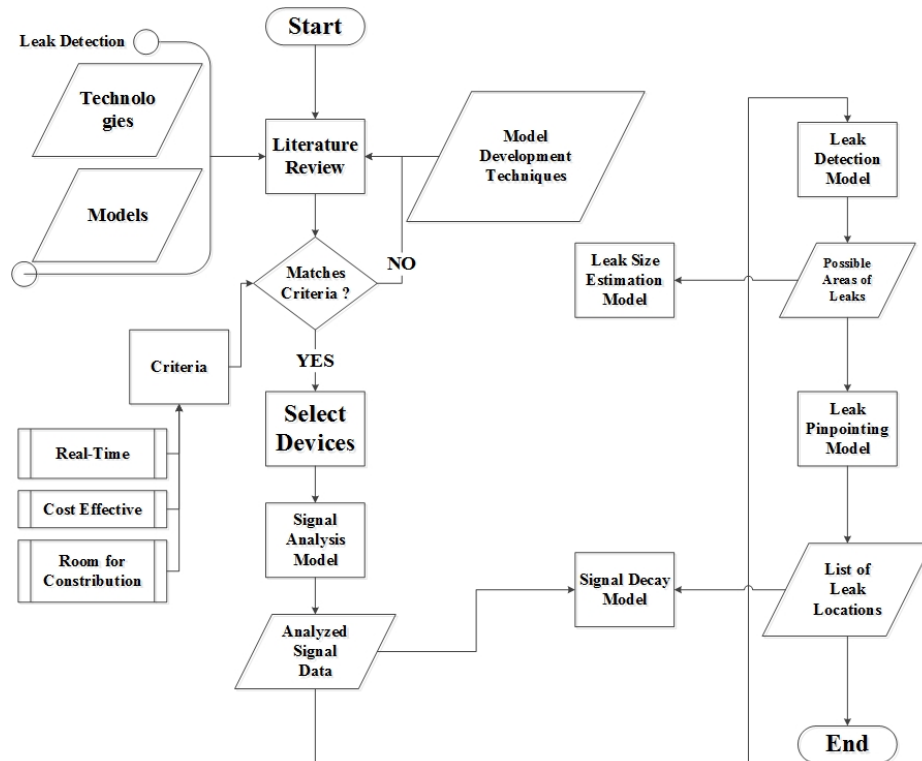


Figure III-1: Accelerometer Real-Time Monitoring System General Methodology

III.1.1 Vibration Signal Analysis Model

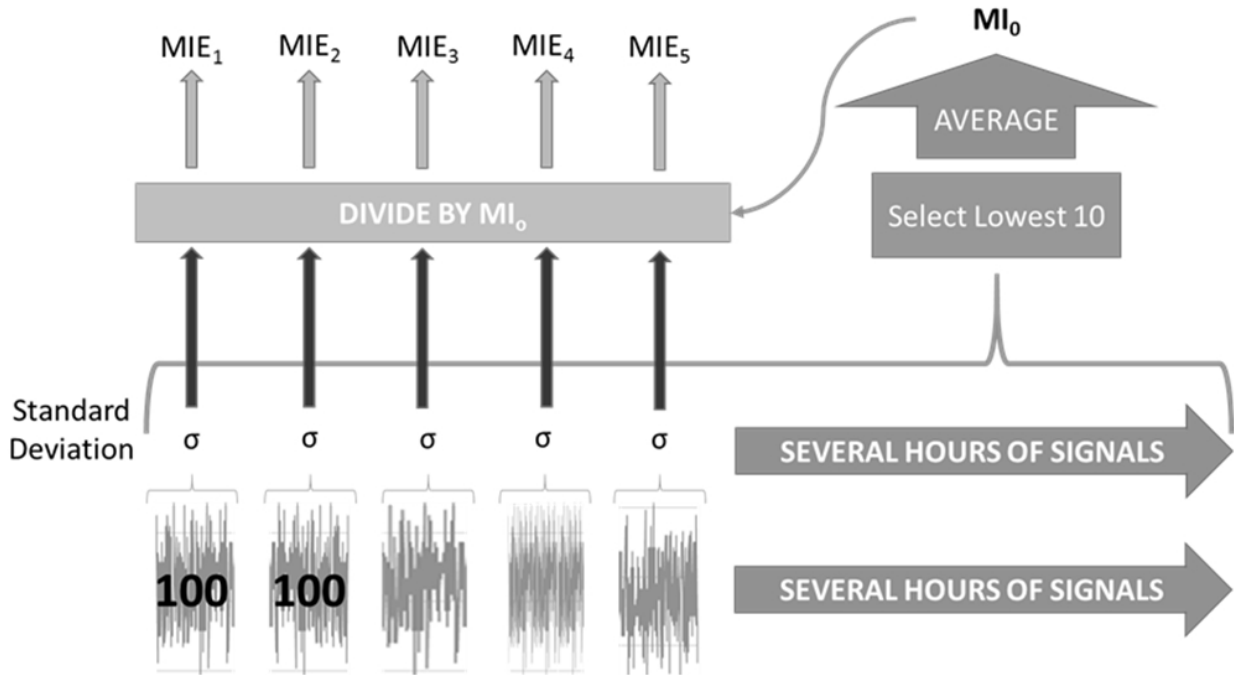


Figure III-2: Vibration Signal Analysis Model

After thoroughly reviewing Martini's model, specifically section II.1.5, the previous diagram in Figure III-2 illustrates the flow of the process. First, multiple hours of signal data at various points were recorded using the a measuring accelerometer to record the signal. The value of the recorded data for this research is around eleven hours. For every 100 seconds, the standard deviation was determined. During the no-leak state, the lowest ten standard deviations were selected and averaged to determine the baseline monitoring index MI_0 . The previously determined standard deviations were divided by the newly calculated (MI_0) to determine the Monitoring Index Efficiency (MIE), which are later utilized to identify leak states from no-leak states and to pinpoint the leak.

III.1.2 Accelerometer Leak Detection Model

The goal of the leak detection model is to instantly make the user alert to the existence of a leak in a particular segment. Thus, to develop this model, the first task is to perform multiple experiments mainly on PVC and ductile iron pipelines with the sizes of one-inch and two-inch, as in Figure III-4. Afterward, using the signal received from the accelerometers in simulated leak and no-leak states and their respective collected MIE, the model inputs are organized. The form of the model input is represented by the predetermined MIE and the relevant state, for example (MIE = 0.982, State = No Leak). The overall approach for the development of this study can be summarized in Figure III-3. The first step is to explore the literature, and study the available classification techniques and explore the current advancements in leak detection using accelerometers. The second step is to set up a series of relevant experiments to explore the interaction of leaks and accelerometers. Using the collected data from the experiments, the models will be developed using the techniques: (1) Linear SVM, (2) Decision Tree and (3) Naïve Bayes. After the models are developed, they are cross-validated using the available data sets to determine their accuracy. In case the models were inaccurate, the development process and the experiments are reassessed. Finally, the accuracy and validity of the developed models are determined and the best models are selected. The developed models will be assessed using the three following parameters: Accuracy, class recall and class precision. Class recall can be defined as “the number of correctly classified positive cases divided by the total number of actual positive cases in the dataset.” Class recall is used to calculate the percentage of positive classes that are classified accurately. However, class precision can be defined as “the number of correctly classified positive cases divided by the number of cases that are classified positive by the model.” Class precision is used to calculate the percentage of positive classes accurately

predicted by the total predicted classes in the positive class (Hossin and Sulaiman 2015; Sokolova and Lapalme 2009).

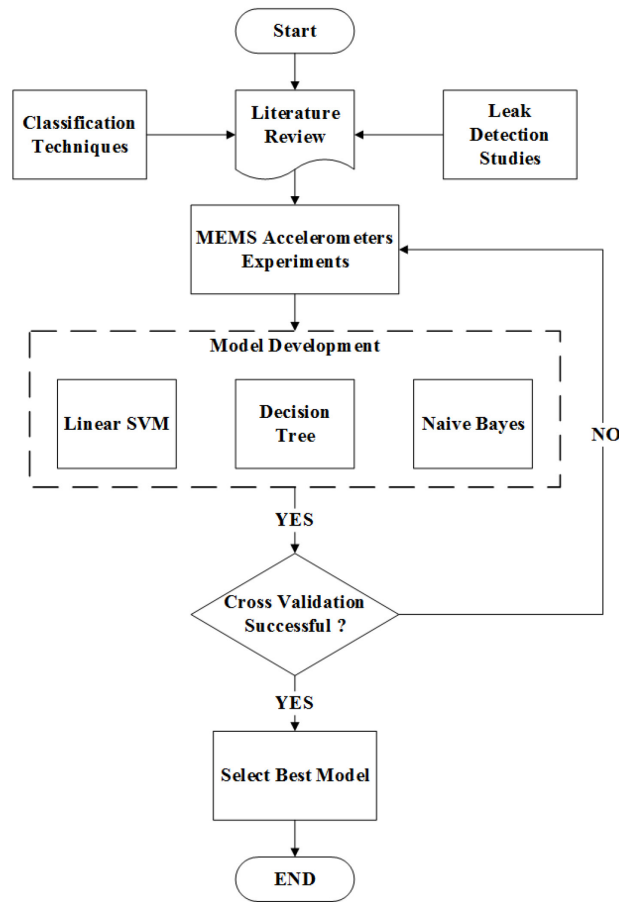


Figure III-3: Leak Identification Model by Vibration Signals

Earlier in this research, the index for assessment (MIE) and the required target as an MIE threshold for identifying various leak states were defined. As in Figure III-4, the next step is to do experiments on PVC and ductile iron pipelines with the sizes of one inch and two inches each. The analyzed data of the experiments are collected and fed into three model development tools using the Rapid Miner 7.4 platform. Each tool develops a model using the acquired data; then the respective models are validated using cross-validation.

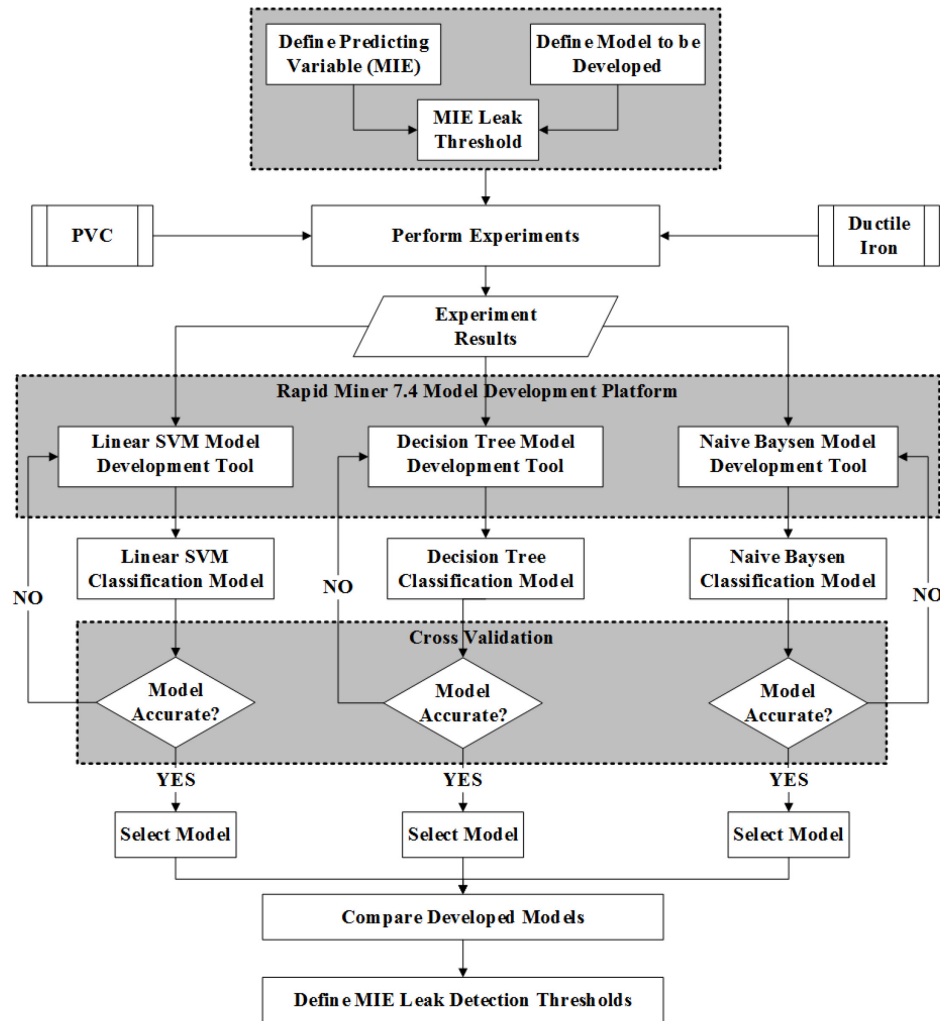


Figure III-4: Accelerometer Leak Identification Model Development Approach

The models that pass cross-validation with the highest accuracy are selected and their results are compared to define the required thresholds. In this research, two types of models are targeted. Leak identification models and leak size classification models. Leak identification models aim to find a particular plane that separates the leak state from the no leak state (existing or non-existing). The input for these models is a data set that contains the value of MIE versus the state of the leak. Leak size classification models go a step further in terms of identify the size of the leak. Those models try to identify small leaks ranging from 10% of the flow rate up to 25% and

big leaks ranging from 25% to 50% of the total flow rate in the pipeline. The assumed ranges are similar in the available markings on the valves used in the experiment.

III.1.3 Accelerometer Leak Size Estimation Model

The development of the leak size estimation model uses the same experimental data collected in section III.1.2. The data similarly is also validated using cross-validation using the rapid miner platform. As in Figure III-5, the main differences are in the way the data was input and the technique the experiment was done with. For the development of this model, instead of two states, there were three states, namely No Leak, Small Leak and Big Leak. A small leak is assumed to be between 1 to 25% of the total flow rate and a big leak is estimated to be 25% to 50% of the total flow rate. SVM will also be the development technique of this model.

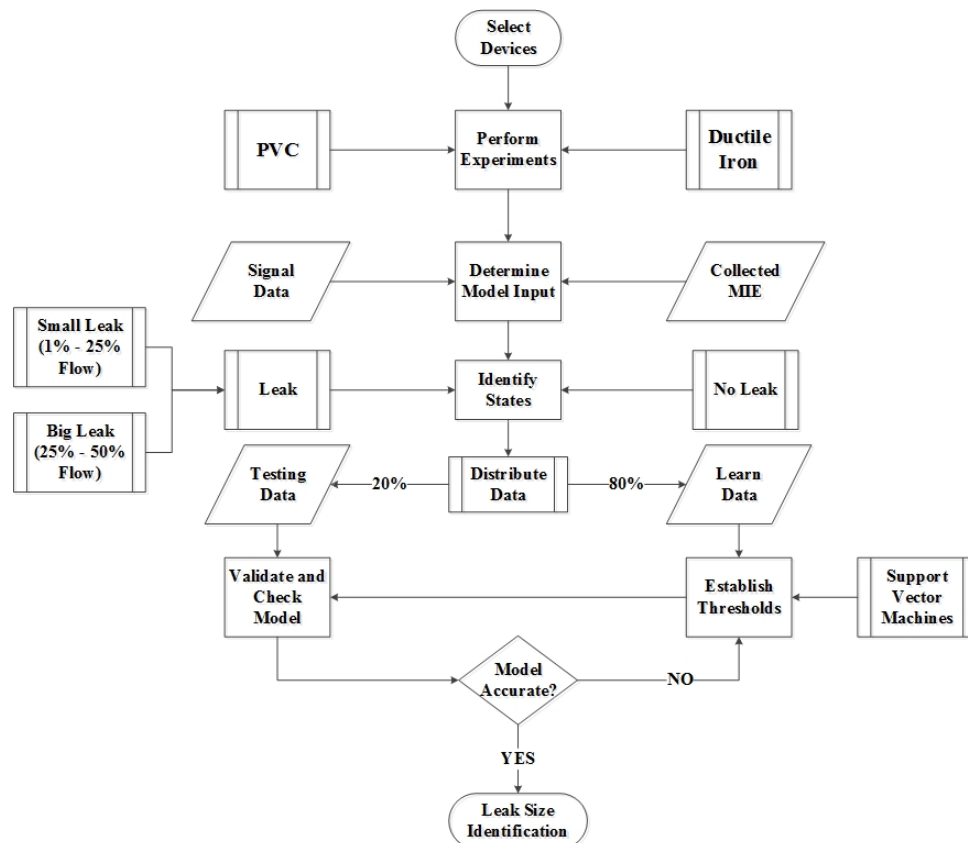


Figure III-5: Leak Size Identification Model by Vibration Signals

III.1.4 Accelerometer Leak Pinpointing Model

After identifying the size of the leak, the next step is to identify the origin of the disruptive signal (i.e. pinpointing the leak). Figure III-6 displays the approach for developing a leak pinpointing model. After device selection and multiple experiments, the signal data was analyzed and the values of the monitoring index efficiencies were determined. The collected data mainly consists of the MIE of the left and right sensors of an expected leak location and the distance between them.

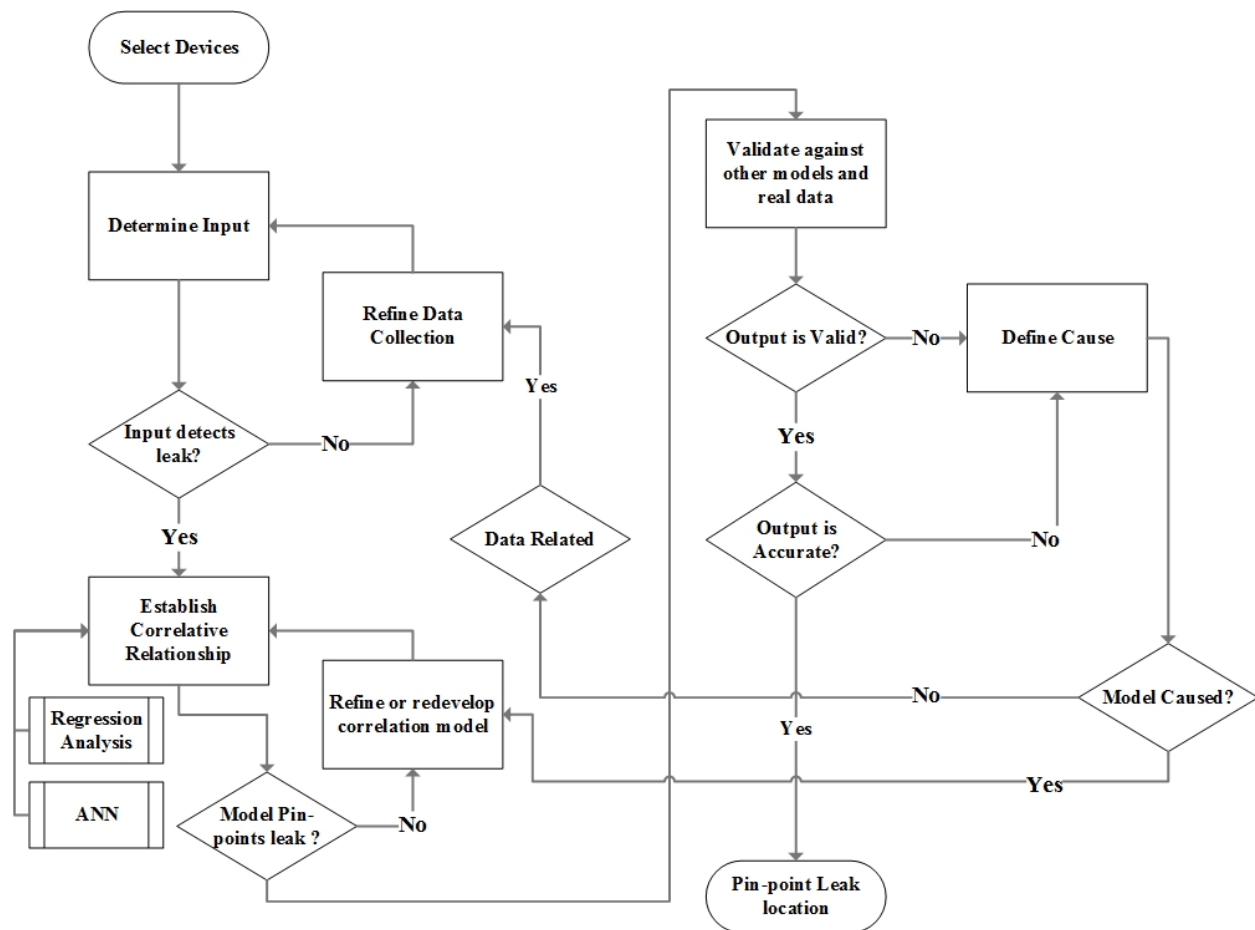


Figure III-6: Leak Pinpointing Model by Vibration Signals

The inputs were a variety of iterations of the collected data such as MIE_T , representing the total value of detected MIE between the left and right sensors and the proportions of left and right over the total. These values also interacted with the distance to detect the relativity amidst the factors. Also, the factors were squared to detect any relation during statistical analysis. To develop the model, artificial neural networks and regression analysis were utilized. The first step in developing the regression model was best subset analysis. Best subset analysis was used to filter out the least relevant parameters previously developed and to identify the most relevant factors helpful in terms of developing an efficient model. The newly selected relevant factors will then be utilized into Minitab and a multiple step regression has been performed to develop a valid leak pinpointing equation. Analysis of Variance (ANOVA) will be performed on each developed model and the model with the best ANOVA results will be selected. When it comes to ANN, the factors determined in the best subset analysis will be used to select the inputs for ANN. The output of ANN will be then compared to the results of regression analysis, and afterward, a sensitivity analysis of the inputs will be performed. This model aims to pinpoint the leak based on immediate data transmission within a 25-centimeter radius from the real leak location.

III.1.5 Vibration Signal Decay Model

After analyzing multiple hours of signal data and viewing the relative MIE values of sensors near and far from the leak, the decay in the amplitude of vibration signals was discovered. To understand the decay, the model illustrated in Figure III-7 was developed. The figure shows that, to assess the dissipation of the signal, two central assumptions were made. The first is that the steady-state baseline affects signal detection by its minimal variation; and the second is that the

steady-state baseline's variation is too negligible to create a significant impact on the motion of the signal. When the experimental data on both PVC and ductile iron are used, the two mathematical theories can be put to the test to allow for the comprehension of the leak decay phenomenon and to understand the changes that happen a signal at a certain distance. The experiments aim to create a baseline at the origin of the leak and then assess the measure of the signal at farther locations. The development of leak impact assessment is based on the uniqueness of the baseline of each sensor that is affected by a variety of factors such as manufacturing flaws, location and external noise. Therefore, it is paramount to assess if the baseline is impactful in signal decay.

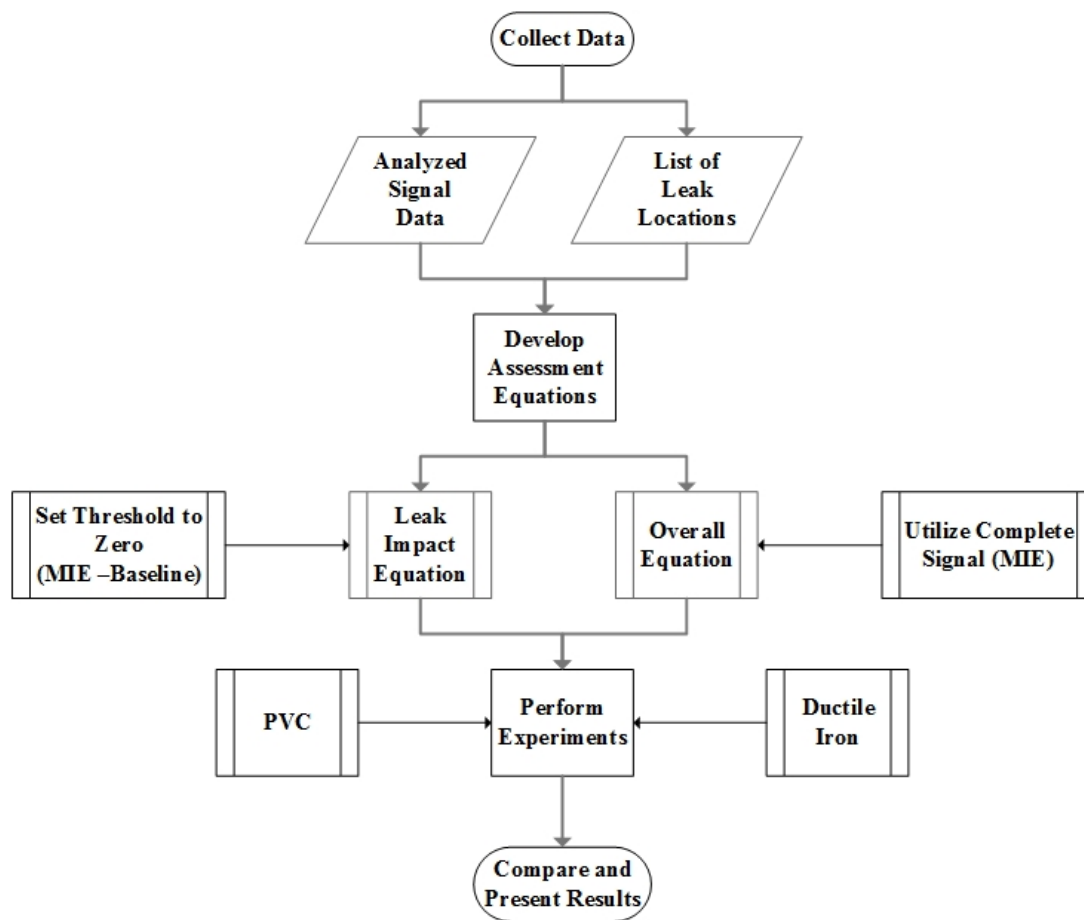


Figure III-7: Vibration Signal Decay Model

As a result, the following experiment is shown in Figure III-8 and Equations (III-1) and (III-2) are developed. Equation (III-1) relies on measuring the MIE value at the leak location, which is sensor 2 in Figure III-8. The main assumption of the equation is that MIE can fully represent the leak signal. The values at remote sensors, i.e. sensors 1 and 3 in Figure III-8, are then measured and are dubbed MIE_{sensor} in Equation (III-1). The values at the sensors are subtracted from the value determined by the sensor at the leak location. The result of the subtraction is then divided by the distance between the two sensors. Equation (III-2) uses the same concept as Equation (III-1). In contrast Equation (III-1), Equation (III-2) assumes that the baseline MI_0 must be removed such that the remaining MIE represents the leak signal. This allows to assess how the leak impacts the reception of the signal at each sensor.

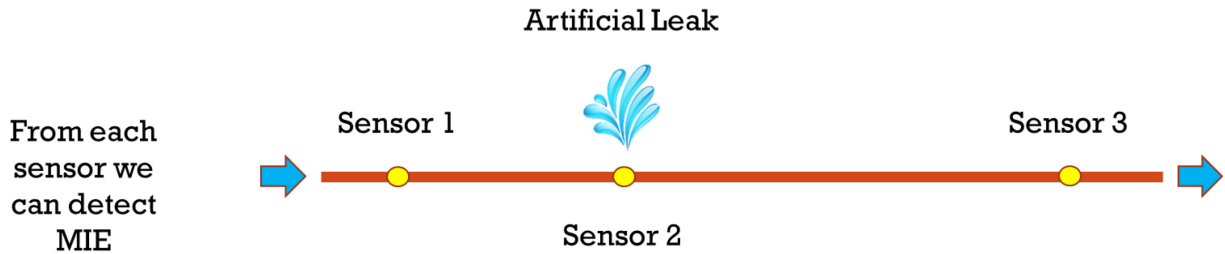


Figure III-8: Decay Theory Experiment Setup

The equations developed for the assessment of the phenomenon are as follows:

$$MIE \text{ Decay: } c = \frac{MIE_{\text{leak}} - MIE_{\text{sensor}}}{D} \quad (\text{III-1})$$

$$Leak \text{ Impact Decay: } c = \frac{LI_{\text{leak}} - LI_{\text{sensor}}}{D} \quad (\text{III-2})$$

Where:

- c = the decay of the monitoring index efficiency per unit meter (m^{-1}).
- MIE_{leak} = the value of the monitoring index efficiency at the leak when the leak occurs.

- MIE_{sensor} = the value of the monitoring index efficiency of the remote sensor under study.
- D = the distance between the sensor at the leak location and the sensor under study (m).
- $LI_{\text{leak}} = MIE_{\text{leak}} - MI_0$ = the exact value of the leak impact at the initial leak location calculated by subtracting the baseline average MI_0 at the no leak state.
- $LI_{\text{sensor}} = MIE_{\text{sensor}} - MI_0$ = the exact value of the leak impact at the remote sensor under study by subtracting the baseline MI_0 of the sensor at the no leak state.

III.2 Data Collection for Accelerometers

This section reviews experiments done to collect signal data from accelerometers in the pipelines subject to leaks. They are conducted in Qatar in partnership with Qatar Foundation and the University of Qatar. In these experiments, four pipelines of two material types are used: (1) PVC and (2) Ductile Iron. Besides, of each material type, two sizes are used: (1) One-inch and (2) Two-inch. The collected data amounts to eight hours of streaming signal at various states.

III.2.1 Experimental Work

The following experiments are designed to study the possible potential of accelerometers in a leak detection setting. Therefore, they present only the necessary conditions for experimentation which are a pressurized pipeline with flowing water and induced leaks. In reality, leaks grow over time and therefore, their progress is not as instantaneous as in those experiments. Additionally, pipelines in the real world operate in confined spaces such as underground and walls, and they have branches and extensions and valves. The aforementioned missing factors are of interest in future research but currently the goal is to realize the pattern of leak vibration signals.

To understand the reaction of accelerometers in the presence of leaks, multiple experiments are required on multiple levels. The first level is to understand the readings noted by the MEMS devices. Thus, at this stage, the first step is to setup any pipeline and place multiple sensors over the pipeline. Figure III-9 displays the general setup of all experimental pipelines of all material types along with the relative distances and sensor placements. Each point noted as P(n) is a valve that can be opened at multiple values to simulate the leak. Pressurized water is inserted through point P1 and exits through point P7. Sensor locations are identified by the term S(n). The locations of the sensors are varied in the experiments to have a more complete view of the problem under study. The image in Figure III-10 displays a two-inch ductile iron pipeline supported by two concrete blocks before the installation of the accelerometers and testing. The valves will act as leak simulators as they are slowly opened and closed.

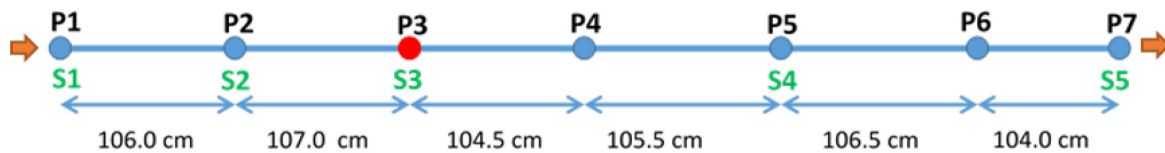


Figure III-9: Experimentation Setup Diagram

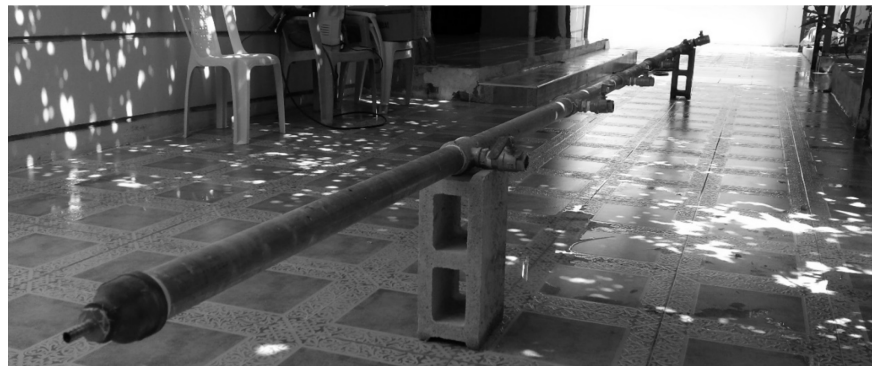


Figure III-10: Two-inch Ductile Iron Pipeline

During experimentation, it was observed that the existence of an immediate open valve at the end creates strong signals that may propagate throughout the whole body of the pipeline and disrupt

the data collection process. Thus, the solution has been to create a damper that would allow water to flow outside of the pipeline without creating a violent vibration that would disrupt the experiment. Figure III-11 shows the solution in the form of a hose extension connected to the exit.

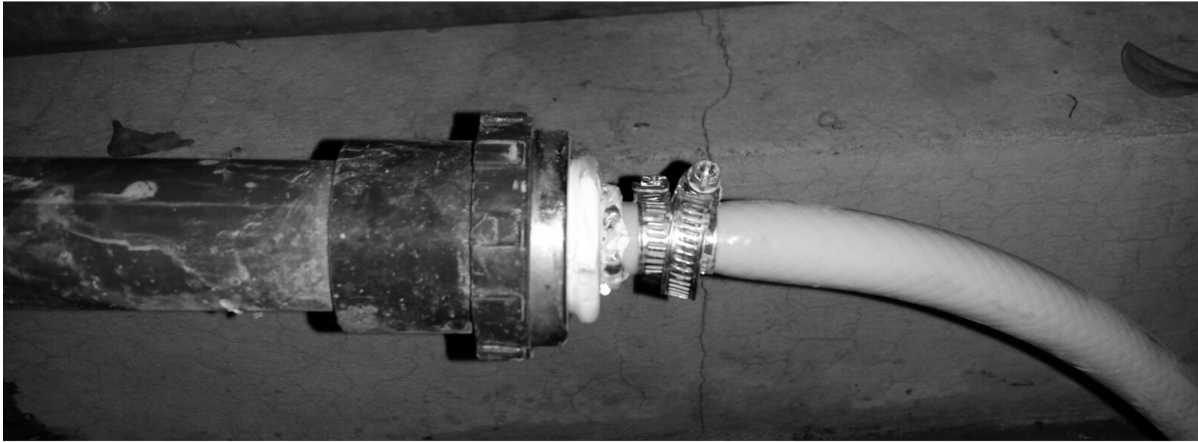


Figure III-11: Pipeline Exit with Release Extension

Figure III-12 displays how the sensors would be placed on each valve following the general installation of the pipeline and before the initiation of the water flow. Once water begins to flow within the pipeline, it takes a certain amount of time to reach a pressurized state and maintain a uniform flow. Once the inlets and outlets have a cohesive uniform flow, the experiments can start. The figure on the right also shows one of the performed experiments on a one-inch PVC pipeline, where a small leak is simulated via having a small opening in the valve.



Figure III-12: One-inch and Two-inch PVC Pipelines

The conducted experiments can be further illustrated in Figure III-13. The figure displays the time distribution of opening and closing the water valves that simulate the leak for two types of experiments: the leak identification experiments, aiming to create a threshold between the leak state and the no-leak state, and the leak size experiments, aiming to distinguish between various leak sizes. The leak identification experiment was conducted, as shown in the figure, by allowing the water into the pipeline without inducing any leaks for the first five minutes. Then, the next five minutes the valve will open and a leak will be induced. After another five minutes pass, the leak would be closed and five minutes for the no-leak state will be induced. The process would be repeated for one hour. The other form of experiment is the leak size experiments. This set of experiments includes similar features to the leak identification experiments, yet leaks are defined more complexly. Small leaks are assumed to have a flow between 1% to 25% of the flow rate in the pipeline, whereas big leaks are assumed to have a flow rate of 26% up to 50% of the overall flow rate. As Figure III-13 shows, leak size experiments are conducted by allowing the pressurized water to flow with no leaks for five minutes, then the leak size is increased to small leaks and another set of five minutes are allowed to pass. Afterwards, the leak size is increased to

the level of a big leak for the next five minutes, and finally, the leak is closed and five minutes of no leaks would be conducted. The process is then repeated for one hour.

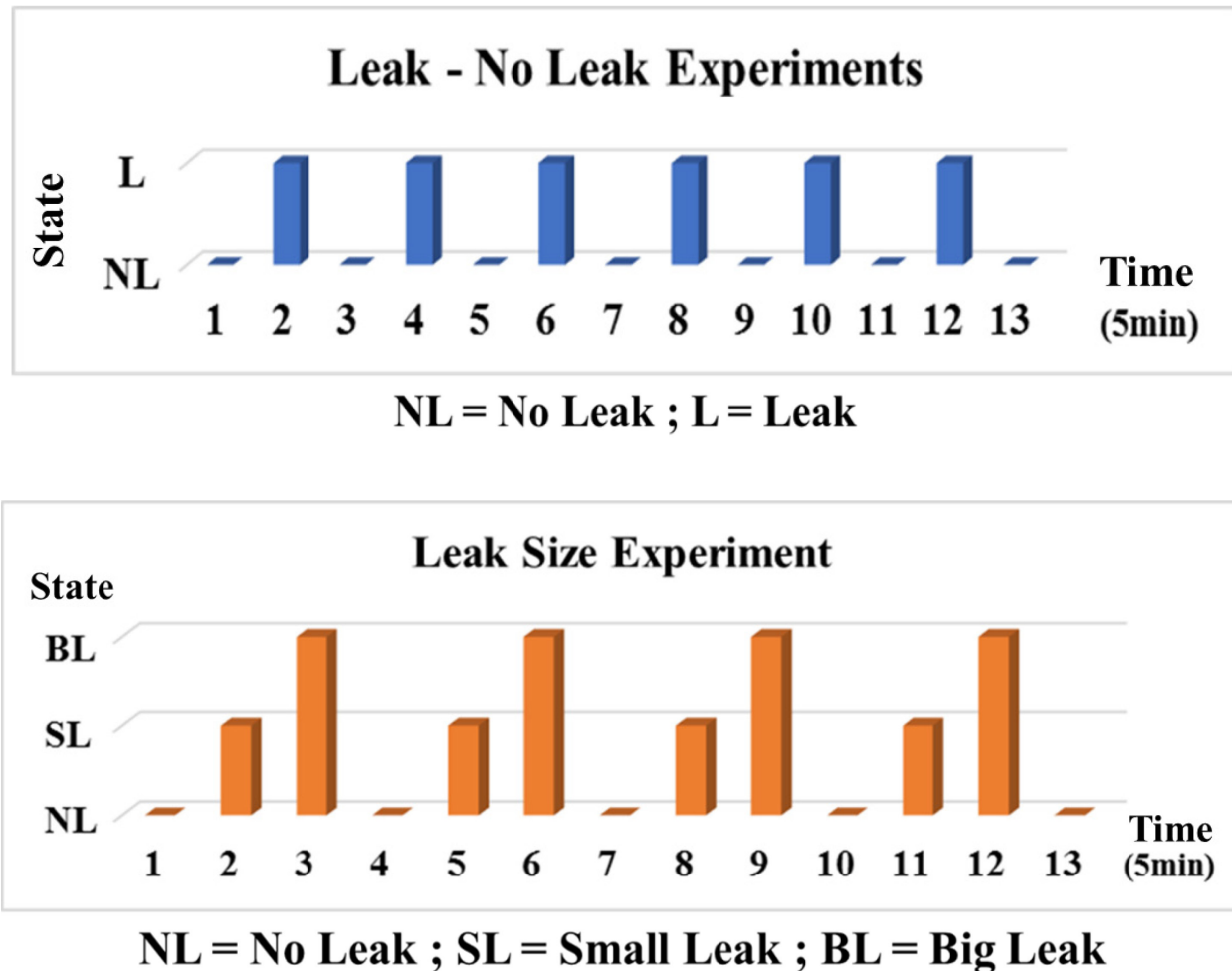


Figure III-13: Time Distribution of Accelerometer Experiments

III.2.2 Data collection

On the level of data collection, the used accelerometers provide measurements of vibration signals in gravitational force units (g). Each second, these values are recorded for each accelerometer to monitor and identify the difference between states. Figure III-14 displays the

data sent by Sensor 1 during a one-hour on-off experiment. An on-off experiment is an experiment where the pipeline is allowed to behave normally with no leaks for the first five minutes and then a leak is induced under the sensor for five minutes and is turned off again then on again till the timer reaches one hour. The collected data is vibration signals and it is measured in (g) as time progresses until one hour has elapsed. Each second represents one data point as data is collected at each second. The figure shows that the duration with no leak tends to be more stable and cohesive. On the contrary, when leaks are induced, the signal would become highly unstable and turbulent, in addition to being much more prominent in terms of value.

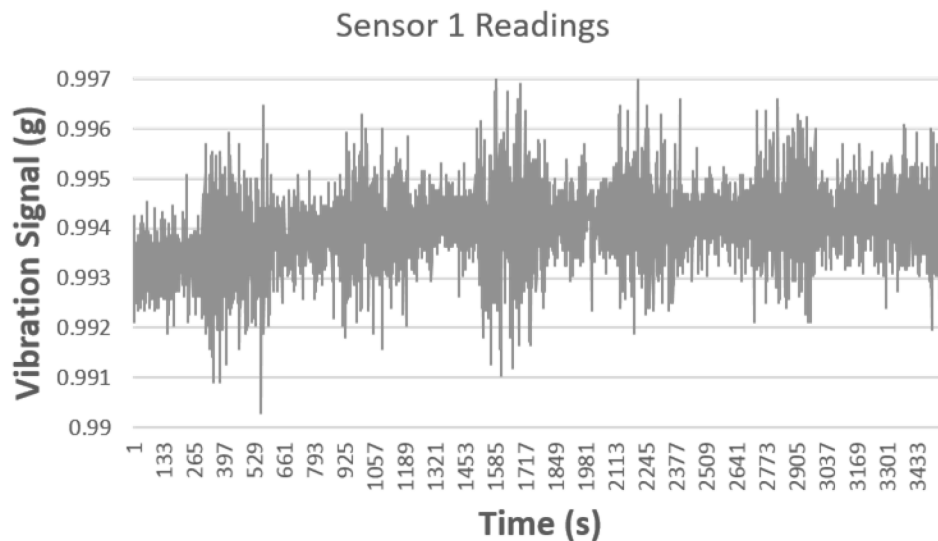


Figure III-14: Sample of Received Data during Experimentation

Using the previously mentioned model in 2.3, developed by Martini et al. in 2016, a Matlab code was developed. The code is designed to automate the process of data collection from sensors, then necessary calculations are done to decipher the received signal data and present incomprehensible monitoring index efficiency (MIE) values. The software takes in data sets of second by second readings and analyzes them per model specifications to develop a bar chart

that shows us the condition of the pipeline by means of MIE for time span $t=100$ seconds. Figure III-15 shows the result of the analysis of the signal received from sensor 1, that is illustrated in Figure III-14. The figure shows a clear distinction between the leak and no leak states within the signal, by measuring the MIE of the signal for each 100-second span until one hour has elapsed. The leak signal displayed in black has a much higher monitoring index efficiency value than that of the no leak signal displayed in grey. And therefore, we can deduce that leaks create more turbulent signals within the system.

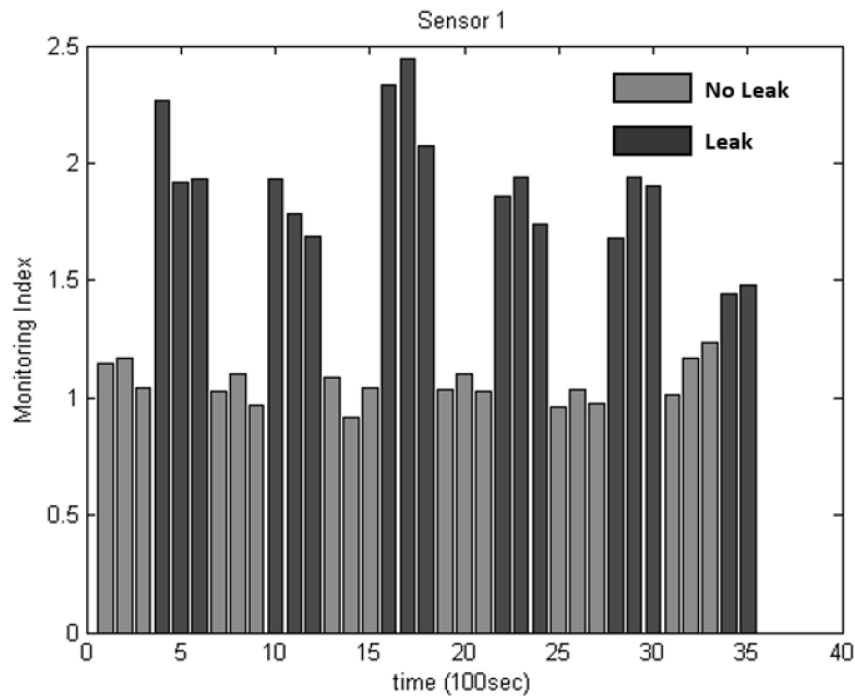


Figure III-15: Sample Matlab Analysis of Sensor Signal

III.2.3 Developed Indicators

Based on the aforementioned data-collection process, multiple indicators were developed in each experiment. The indicators and variables are aimed to give a representation for the states of the pipeline under study. The indicators and variables are defined and illustrated in this section.

Free Variables

Free variables are the intended and designated output under study and the developed models aim at identifying free variables. In this part, two free variables are available and they are X_L and X_R .

Where:

- X_L is the distance from the sensor on the left of the expected leak to the location of the leak.
- X_R is the distance from the sensor on the right of the expected leak to the location of the leak.

Indicators

Indicators are values that represent the current state of the pipeline between two sensors at any given time. Some indicators are fixed, such as the distance between the sensors and its derivatives. The indicators in model development are as follows.

- D is the total distance between the two sensors on both ends of the leak and the summation of X_L and X_R must be equal to D after both are determined.
- MIE_L is the monitoring index efficiency of the signals detected by the sensor to the left of the suspected leak.
- MIE_R is the monitoring index efficiency of the signals detected by the sensor to the right of the suspected leak.
- MIE_T is the total monitoring index efficiency and it is equal to the sum of the left and right monitoring index efficiencies, MIE_L and MIE_R .
- L/R is the relative value of MIE_L over MIE_R .
- R/L is the relative value of MIE_R over MIE_L .

- L/T and R/T are the relative values of MIE_L and MIE_R over the total monitoring index efficiency MIE_T .
- LD/T and RD/T are multiplications of the relative monitoring efficiencies, MIE_L and MIE_R , over the total monitoring index efficiency, MIE_T , with the total distance, D.
- $(L/T)^2$, $(R/T)^2$ and D^2 are the squares of the initially identified values.

Those values are determined by experiments and organized in an excel sheet for all the performed experiments on both PVC and ductile iron pipelines with the sizes of one inch and two inches. Figure III-16 illustrates a sample of the excel sheet of the collected experiment-driven data. This data will be later used in the development of leak detection models using regression analysis, support vector machines and artificial neural networks. The factors enlisted in the figure below is explained earlier in this section.

MEIL	MEIR	MEIT	L/R	L/T	R/T	XL	XR	D	R/L	D ²	(L/T)*D	(L/T) ²	(R/T)*D	(R/T) ²
24.5	304	328.5	0.08059	0.07458	0.92542	-2.37	0	2.37	12.4082	5.617	0.17676	0.006	2.193	0.856
24.5	72	96.5	0.34028	0.25389	0.74611	-2.37	2.56	4.93	2.93878	24.305	1.25166	0.064	3.678	0.557
123	304	427	0.40461	0.28806	0.71194	-1.21	0	1.21	2.47154	1.464	0.34855	0.083	0.861	0.507
123	213	336	0.57746	0.36607	0.63393	-1.21	1.26	2.47	1.73171	6.101	0.9042	0.134	1.566	0.402
123	72	195	1.70833	0.63077	0.36923	-1.21	2.47	3.68	0.58537	13.542	2.32123	0.398	1.359	0.136
304	213	517	1.42723	0.58801	0.41199	0	1.26	1.26	0.70066	1.588	0.74089	0.346	0.519	0.17
304	72	376	4.22222	0.80851	0.19149	0	2.47	2.47	0.23684	6.101	1.99702	0.654	0.473	0.037
7	41	48	0.17073	0.14583	0.85417	-2.13	0	2.13	5.85714	4.537	0.31063	0.021	1.819	0.73

Figure III-16: Sample of the Collected Data Sheet

III.3 Results of Developed Models

For leak detection, two models are currently available. An SVM model is developed to differentiate between leak states and identify the existence of a leak. For leak pinpointing, a set of two equations is developed to identify the exact location of a leak using regression analysis. The leak pinpointing model is general and can be utilized for both PVC and ductile iron pipelines of sizes one and two inches.

III.3.1 Leak Detection and Size Identification Models

The initial step in the development of an accelerometer-based leak monitoring system is to identify the existence of leaks with leak detection models. This section presents the implementation of the developed models for leak detection and leak size identification. The first step would be to identify a threshold for leaks using vibration signals. The second step would be to identify the significance of the leak through identifying its size.

III.3.1.1 Leak Identification Model Implementation

For leak state identification, the primary assessment criterion is the capability of the devices and the models to find a separating plane between the leak state and the no-leak state. After cross-validating each model, the results are collected and displayed in this section and detection thresholds for each model are identified.

Linear SVM Model

For SVM, the model presents a threshold of $MIE = 1.018$ in Equation III-3. The model is cross-validated against the original 282 MIE data points and the results are summarized in the first row of Table III-1. Table III-1 shows that the accuracy of the Linear SVM model is averaged at 96.44% with an average deviation of 5.76% increasing or decreasing. Table III-1 also shows that the Linear SVM Model has missed nine leaks and categorized them under no-leak states. Furthermore, the model is capable of identifying and retrieving No Leak and Leak data accurately. So, the class-recall values are 93.31% and 93.38% for No Leak and Leak states respectively. Additionally, the quality of the retrieved data per class is high as the class precisions are 94.12% for the No Leak state and 99.22% for the Leak state.

$$if\ MIE \begin{cases} \leq 1.018, & No\ Leak \\ > 1.018, & Leak \end{cases} \quad (III-3)$$

Decision Tree Model

The Decision Tree Model suggests the value of $MIE = 1.052$ be the threshold, as in Figure III-17 and Equation III-4. The figure also illustrates that any value equal or less than 1.052 is considered a no-leak state with 100% confidence, while any value higher than 1.052 is considered to be a leak. The cross-validation results of this model are summarized in the second row of Table III-1.

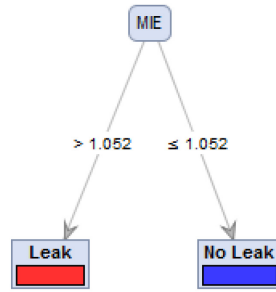


Figure III-17: Decision Tree Leak Identification Model

The accuracy of the model is calculated to be at 99.29% with an average deviation of 1.43%. The model does not classify any leaks as no leaks, while the model classifies two no-leak states as leaks. The model has the highest percentage of leak data retrieval among the developed models with a 100% class recall and 98.55% of the collected data is precisely identified as leaks. As for the No Leak state, the model is capable of accurately collecting No Leak data with a 98.62% class recall and it classifies the collected data into the No Leak state with a 100% class precision.

$$if\ MIE \begin{cases} \leq 1.052, & No\ Leak \\ > 1.052, & Leak \end{cases} \quad (III-4)$$

Naïve Bayes Model

As for the Naïve Bayes Model, Figure III-18 and Equation III-5 show that the separating point between the leak and no-leak states is estimated to be at the intersection point between the two curves where $MIE = 1.07$. The figure displays the densities of each category compared to MIE and therefore the distribution of each category on the MIE axis. The capabilities of the model are validated using cross-validation and the results are summarized in the third row of Table III-1. Table III-1 shows that the average accuracy of the NB model is estimated to be 98.57% with an average deviation of 2.37%.

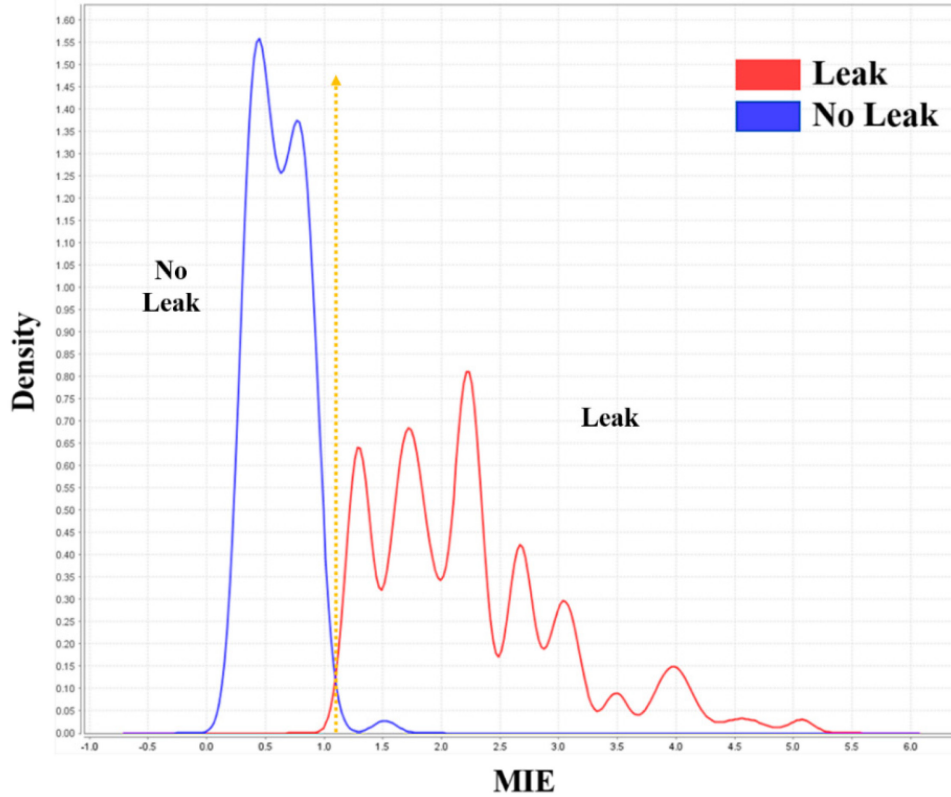


Figure III-18: Naïve Bayes Leak Identification Model

The NB Model has three missed leaks and one false alarm during testing and validation. It has also shown high levels of proper data retrieval with 99.31% and 97.79% class recall for No Leak and Leak states respectively. Additionally, the collected data is classified with high precision. The class precision for the No Leak state is 97.96%, whereas the class precision for the Leak state is 99.25%.

$$if\ MIE \begin{cases} \leq 1.07, & No\ Leak \\ > 1.07, & Leak \end{cases} \quad (III-5)$$

Table III-1: Leak Detection Model Cross Validation Results

Model	Results			
Support Vector Machines	Accuracy = 96.44% +/- 5.76 %	True No Leak	True Leak	Class Precision (%)
	Predicted as No Leak	144	9	94.12
	Predicted as Leak	1	127	99.22
	Class Recall (%)	93.31	93.38	
Decision Tree	Accuracy = 99.29% +/- 1.43%	True No Leak	True Leak	Class Precision (%)
	Predicted as No Leak	143	0	100
	Predicted as Leak	2	136	98.55
	Class Recall (%)	98.62	100	
Naïve Bayes	Accuracy = 98.57% +/- 2.37%	True No Leak	True Leak	Class Precision (%)
	Predicted as No Leak	144	3	97.96
	Predicted as Leak	1	133	99.25
	Class Recall (%)	99.31	97.79	

III.3.1.2 Leak Size Identification Model Implementation

Regarding leak size identification, the models are expected to identify planes that separate between three different states, i.e. (1) No-Leak, (2) Small Leak and (3) Big Leak. Besides, to reassess the consistency of the leak detection models, the data of no-leak states is used in the

development of leak size classification models. A small leak is assumed to have a discharging leak between 10% to 25% of the overall flow rate, whereas a big leak has a discharge rate of 26% and up to 50% of the overall flow rate.

Linear SVM Model

For leak size classification, the Linear SVM Model provides a threshold of 1.018, in Equation III-6, to separate small leaks from no leak condition. The provided value is consistent with the previously developed model. The model also provides the value of MIE = 2.24 to distinguish between small leaks and big leaks. The average accuracy of the model is 80.06% with an average deviation in accuracy, equal to 7.55% as displayed in the first row of Table III-2. The model has retained its capacity to accurately retrieve and identify No Leak data with a class recall of 100% and class precision of 84.8%. For big leaks, the model has an equal class recall and precision amounting to 80%. However, the model has encountered issues in classifying small leaks; the model often misinterprets them as no leaks or big leaks and thus has a low class-recall for small leaks equal to 32.79% and a low class-precision of 57.14%.

$$if \ MIE \begin{cases} \leq 1.018, \text{ No Leak} \\ \in]1.018, 2.24], \text{ Small Leak} \\ > 2.24, \text{ Leak} \end{cases} \quad (III-6)$$

Decision Tree Model

The Decision Tree Model has remained consistent in separating leak states from no-leak states by retaining the threshold of MIE = 1.052, as in Figure III-19 and Equation III-7. Additionally, the Decision Tree Model specifies the value of 1.595 for MIE to represent the threshold separating small leaks from big leaks. If the value of MIE is less than or equal to 1.595, yet

above 1.052, the leak is, with a very high degree of confidence, a small leak. Above an MIE of 1.595, the leak is most probably a big leak while small leaks are also very likely at this range.

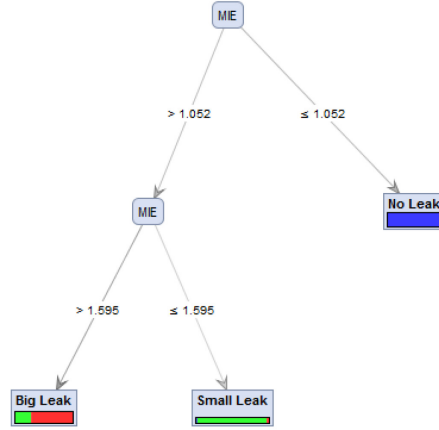


Figure III-19: Decision Tree Leak Size Identification Model

The second row of Table III-2 summarizes the results of the cross-validation of the model. The model has an accuracy of 85.39% with an average deviation of 3.02%. The model has had 2 false alarms and two small leaks that have gone undetected. Furthermore, the Decision Tree Model behaves better than the SVM Model with the class precision and the class recall both equal to 98.62% for the No Leak state plus a class recall rate of 49.18% for small leaks and a class precision of 75%. For big leaks, the class recall shows an improvement in the Decision Tree Model, compared to the SVM Model by 9.33%, whereas the class precision is 10% less than in the SVM Model.

$$\text{if } MIE \begin{cases} \leq 1.052, \text{ No Leak} \\ \in]1.052, 1.595], \text{ Small Leak} \\ > 1.595, \text{ Leak} \end{cases} \quad (\text{III-7})$$

Naïve Bayes Model

As for Naïve Bayes, the leak versus no leak separation threshold remains consistently at MIE = 1.07, as in Figure III-20 and Equation III-7. On the other hand, the threshold between a small and

a big leak is estimated to be at the meeting point between the downward curve of the small leak and the upward going cure of the big leak state. Thus, the threshold is the straight-line MIE = 1.88. In terms of accuracy, the third row of Table III-2 shows that the NB Model has an average accuracy of 86.8% with an average deviation of 6.01%. The model has had one false alarm and two small leaks going undetected. The table also shows that the Naïve Bayes Model provides a more stable model than the other techniques with a class recall of 99.31% for the No Leak state and a class precision of 98.63%. The NB Model provides the highest classification for small leaks with a 63.93% class recall and 72.22% class precision. The classification of big leaks using the NB Model is acceptable with an 81.33% class recall and a 75.31% class precision.

$$if\ MIE \begin{cases} \leq 1.07, & \text{No Leak} \\ \in]1.07, 1.88], & \text{Small Leak} \\ > 1.88, & \text{Leak} \end{cases} \quad (III-8)$$

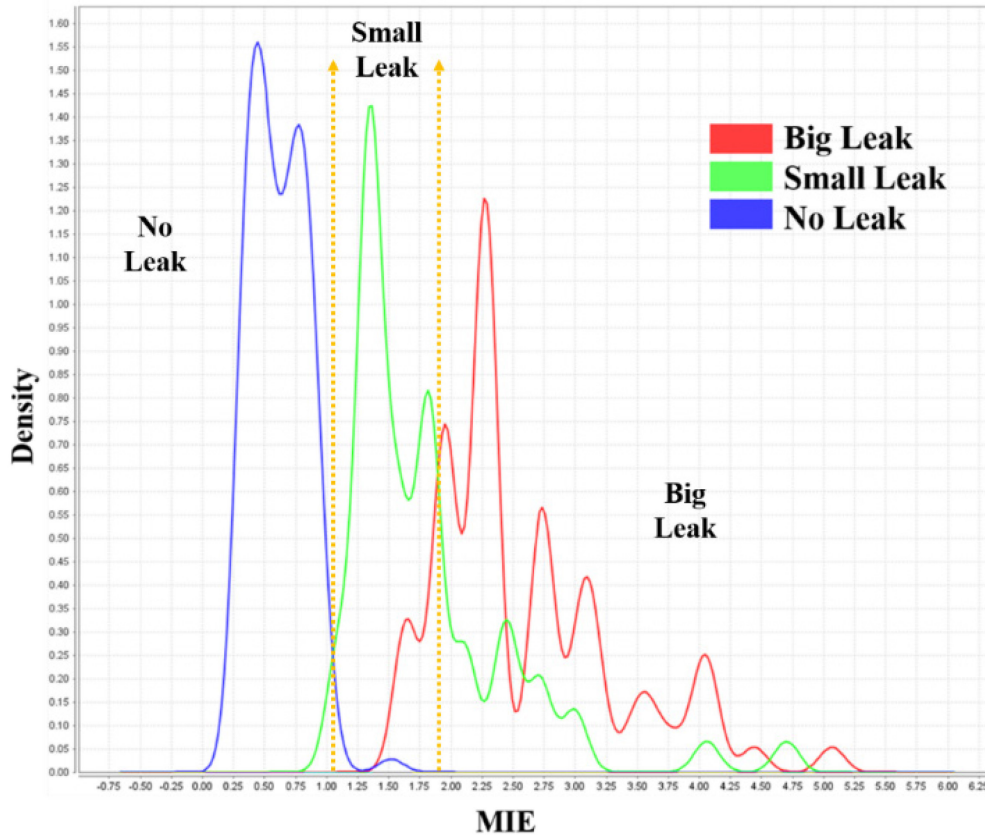


Figure III-20: Naïve Bayes Leak Size Identification Model

Table III-2: Leak Size Identification Model Cross Validation Results

Model	Results				
Support Vector Machines	Accuracy: 80.06% +/- 7.55%	True No Leak	True Small Leak	True Big Leak	Class Precision (%)
	Predicted as No Leak	145	26	0	84.8
	Predicted as Small Leak	0	20	15	57.14
	Predicted as Big Leak	0	15	60	80.00
	Class Recall (%)	100	32.79	80.00	
Decision Tree	Accuracy: 85.39% +/- 3.02%	True No Leak	True Small Leak	True Big Leak	Class Precision (%)
	Predicted as No Leak	143	2	0	98.62
	Predicted as Small Leak	2	30	8	75.00
	Predicted as Big Leak	0	29	67	69.79
	Class Recall (%)	98.62	49.18	89.33	
Naïve Bayes	Accuracy: 86.8% +/- 6.01%	True No Leak	True Small Leak	True Big Leak	Class Precision (%)
	Predicted as No Leak	144	2	0	98.63
	Predicted as Small Leak	1	39	14	72.22
	Predicted as Big Leak	0	20	61	75.31
	Class Recall (%)	99.31	63.93	81.33	

III.3.1.3 Discussion of Results

The results presented here show the capabilities of accelerometers to detect the vibrations caused by leaks and accurately identify them. Regarding leak identification, the three used techniques produce high results as summarized in Figure III-21(a). The box and whisker plot represents the median as a thick line that cuts the box into two partitions, each partition representing a quartile of the distribution of the accuracies presented by the model. Besides, the distance from the periphery of the box to each outer line represents a quartile of the data. For example, the Decision Tree data in Figure III-21(a) shows a median of 99.29% and a maximum possible accuracy of 99.8% and minimum possible accuracy of 97.8%. The illustration also indicates that 75% of the time, the accuracy of Decision Tree Model has range from 98.2% to 99.8%.

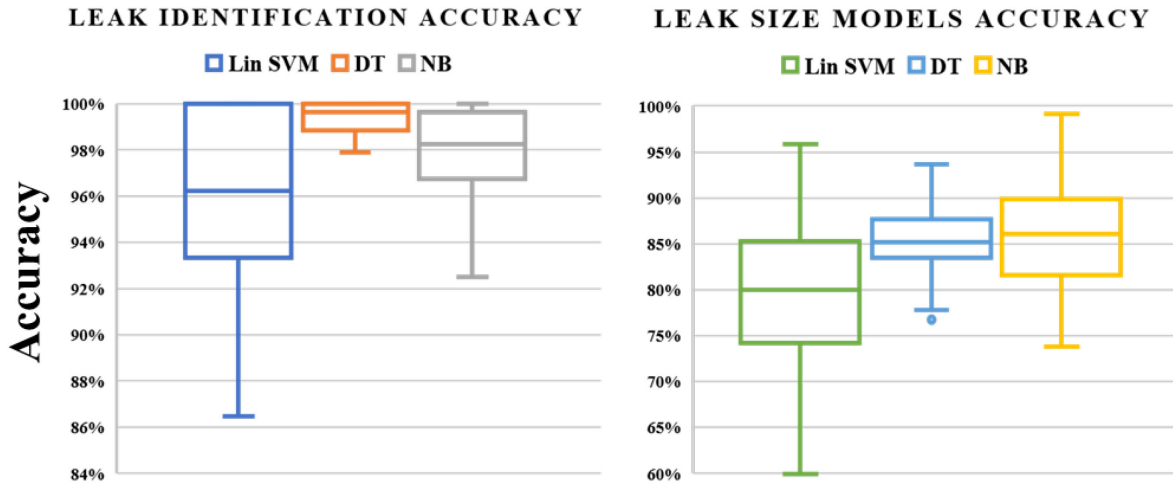


Figure III-21: Box and Whisker Plot for (a) Leak Identification Models and (b) Leak Size Models

Regarding leak identification accuracy, the Decision Tree Algorithm provides the model with the highest accuracy with the minimum deviation at 99.29% accuracy and a deviation of 1.43%. The Decision Tree Model is followed by the Naïve Bayes Model and the Linear SVM Model. For leak size identification, the Naïve Bayes Model provides the highest average accuracy at 86.8% but with a deviation of 6.01% followed by the Decision Tree Model with 85.39% and finally the Linear SVM Model. From the box-and-whisker plot in Figure III-21(b), it can be deduced that the Decision Tree Model has a lesser accuracy than that of the Naïve Bayes Model and that the deviation of the Decision Tree Model is 3.02%, representing a more consistent model development than that of the two other techniques.

The determined thresholds are summarized in Table III-3. The lowest threshold to separate the leak state from the no-leak state is provided by the Linear SVM Model with a value of MIE = 1.018, whereas the highest threshold is provided by the NB Model with an MIE of 1.07. Since the three models show high levels of accuracy, it is possible to view the three thresholds as three

levels of conservativeness with 1.018 being the most conservative threshold and 1.07 being the least conservative threshold separating the leak states from the no-leak state.

Table III-3: Leak Identification Model Thresholds

		ALGORITHM RESULTS (MIE = x)		
		Linear SVM	Decision Tree	Naive Bayes
STATE	No-Leak	1.018	1.052	1.07
	Small Leak	2.24	1.595	1.88
	Big Leak			

Accordingly, the leak size identification thresholds can be assessed the same way, with MIE equal to 1.595 as the most conservative threshold and MIE of 2.24 as the least conservative threshold for separating small leaks from big leaks. Evidentially, the moderately conservative selections for thresholds would be MIE of 1.052 to identify the existence of a leak and an MIE of 1.88 to signify the existence of a big leak. The midpoint solution is considered to be a valid starting point for network assessment.

III.3.2 Accelerometer Leak Pinpointing Model Results

For the leak pinpointing model, the indicators developed in section III.2.3 of this thesis are utilized. The data is collected and input into the Minitab 16 software for analysis. First, a best-subset analysis is performed using the whole data population with X_L and X_R as free variables. Based on the best-subset analysis, the selected factors for the distance from the left sensor to the leak, X_L , are L/R , L/T , D , D^2 and LD/T . As for X_R , the selected factors are R/T , $(R/T)^2$, D , D^2

and LD/T. Based on the preceding, the indicators for each variable are input into Minitab, step-wise regression is performed and a general model is developed.

III.3.2.1 Developed Pinpointing Models

A. General Regression Model

The Minitab results show two equations for calculating the values of the distances from the left and right sensors surrounding the leak. Equation (III-9) displays the relationship developed for determining X_L using the best-subset variables determined via best-subset analysis. In the case of X_L , they are L/R, L/T and D. As one of the main assumptions of this approach, the flow always moves from X_L to X_R . The leak location is also considered an origin point. Based on that, since X_L is against the flow, X_L is considered a negative value with a maximum of zero and X_R is always considered a positive value with a minimum of zero. X_R is illustrated in Equation (III-10) as an interaction between three variables remaining from the best subset analysis and they are R/T, L/T and D.

$$X_L = -2.05 + 0.1718 * \frac{L}{R} + 3.5 * \frac{L}{T} - 0.295 * D + 0.01985 * D^2 - 0.3351 * \frac{L}{T} * D \quad \text{(III-9)}$$

$$X_R = 2.766 - 6.88 * \frac{R}{T} + 2.251 * \left(\frac{R}{T}\right)^2 + 0.4178 * D + 0.0248 * D^2 + 0.3187 * \frac{L}{T} * D \quad \text{(III-10)}$$

Where:

X_L = the distance from the left sensor to the suspected leak.

X_R = the distance from the right sensor to the suspected leak.

L = monitoring index efficiency at the left sensor MIE_L .

R = monitoring index efficiency at the right sensor MIE_R .

T = total monitoring index = L + R.

D = total distance between sensors.

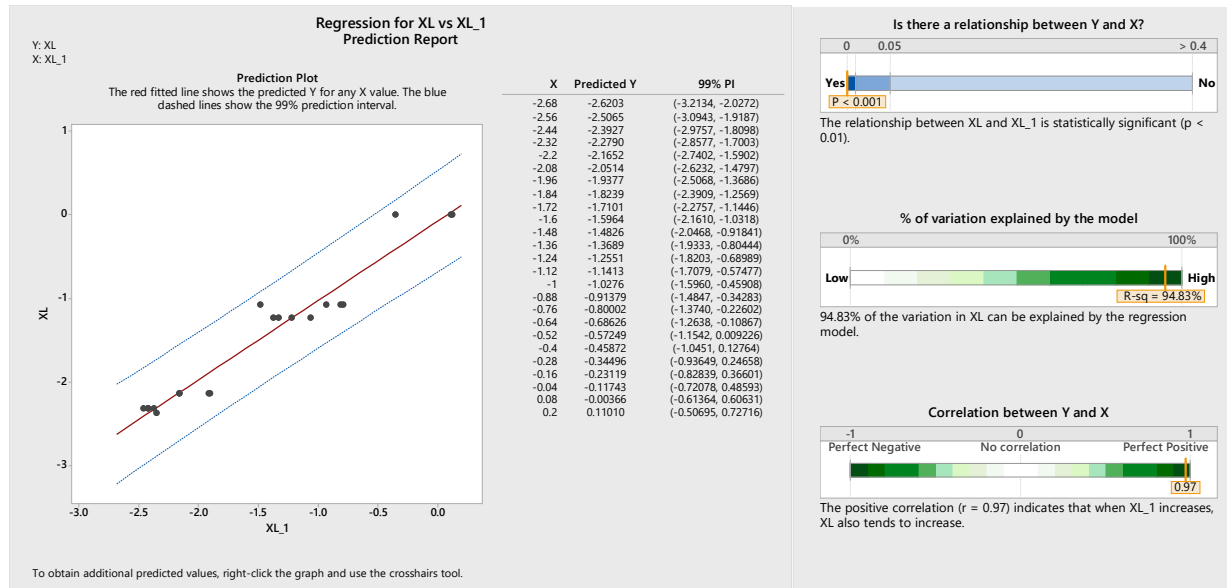


Figure III-22: General X_L Regression Prediction Report

The extent of equation III-9 is studied in Figure III-22 and the first row of Table III-6. The figure shows the testing data used to verify the developed model versus the exact leak locations. The straight line in red represents the developed model and the black points represent the actual leak location during the experiment. The model has proximity to all data points except for two outliers that are relatively far from the model; those high discrepancies can be a consequence of experimental error or other external factors. The figure also shows that the p-value for the model is less than 0.001, implying that the model has a high statistical correlation with the phenomenon under study. Finally, the level of correlation via r-squared is displayed at 94.83%. Besides, Table III-6 shows that the F-test value for equation III-9 is equal to 187, which is a high numerical value. The high value of the F-test result indicates that the developed equation is statistically significant. The table also displays the step p-values for each variable used in the equation. The equation starts off with a zero p-value, yet the p-value increases when D and D^2 are used to be later dropped to less than 0.001 with the implementation of $(L/T)*D$.

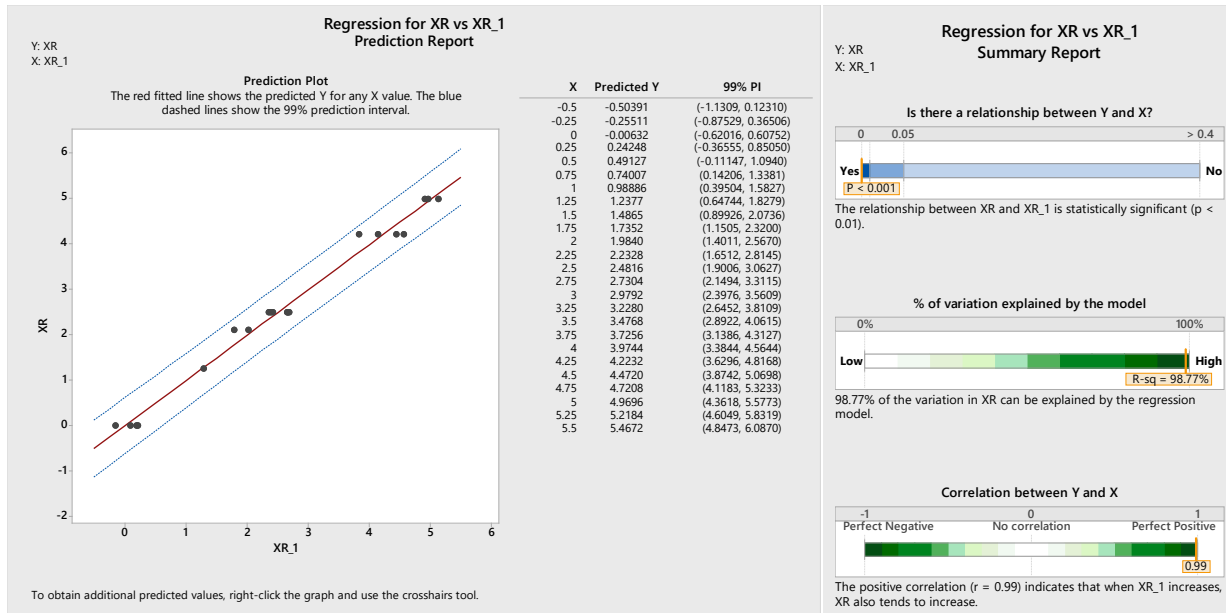


Figure III-23: General X_R Regression Prediction Report

Similarly, a prediction report has been established for the general X_R model, for understanding the behavior of the model as highlighted in Figure III-23. The model displays more concise predictions than its X_L counterpart with no outliers and accurate predictions. The model also displays a high level of correlation through its p-value, which scores a value less than 0.001 to indicate the high significance of the relationship. Furthermore, the model has a high r-squared percentage of 98.77% and a high correlation value between the model and test points of 0.99. Furthermore, the second row of Table III-6 proves that the model is statistically significant given that it has an F-value of 734.5 with a high r-squared percentage close to 100% and a p-value that is less than 0.001. The progression of the p-value displays that the p-value of the model has progressed smoothly with a minimal increase except for D^2 , but the model has readjusted.

B. Ductile Iron Specific Regression Model

Similar to the previously developed general model, a specific model is developed to study the differences between ductile iron pipelines and PVC pipelines. After performing a best-subset analysis, the following models are displayed in Equations III-11 and III-12 for X_L and X_R respectively. X_L relies on MIE_L along with the total MIE, MIE_T and D as the distance between the two sensors, to determine the distance from the leftmost sensor X_L to the leak.

$$X_L = -2.708 + 3.12\left(\frac{L}{T}\right)^2 + 3.194\frac{L}{T} + 0.1134D - 0.754\frac{L}{T}D \quad (\text{III-11})$$

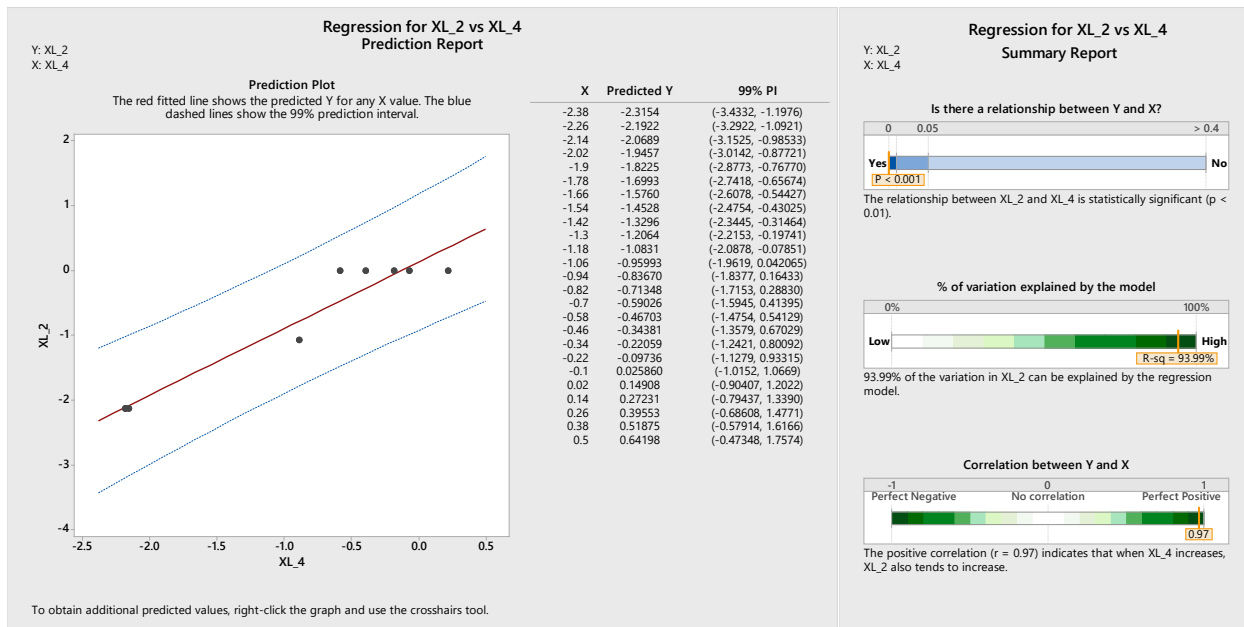


Figure III-24: X_L for Ductile Iron Regression Prediction Report

To delve further into the model, a prediction report is established and displayed in Figure III-24 and the third row of Table III-6. The figure shows the red line that represents the model against the experimental testing data points. The developed model has a small deviation as the figure shows. Also, the model is statically significant because its p-value is less than 0.001 and its r-

squared percentage is at 93.99%. Finally, the correlation factor between the model and the real data is rated at 0.97, showing a high level of representation by the model for the phenomenon. Additionally, the F-value for the developed model is determined at 81.17, which is a value greater than zero and thus attests to the statistical significance of the model. The variables that create spikes in the p-value of the model are L/T and D. Similarly, the model for X_R is displayed in equation III-12 and relies on the value of MIE in the rightmost sensor and the total MIE detected between the two sensors and, finally, the total distance between the two sensors D.

$$X_R = 3.609 - 9.44\frac{R}{T} + 3.12\left(\frac{R}{T}\right)^2 + 0.359D + 0.754\frac{R}{T}D \quad (\text{III-12})$$

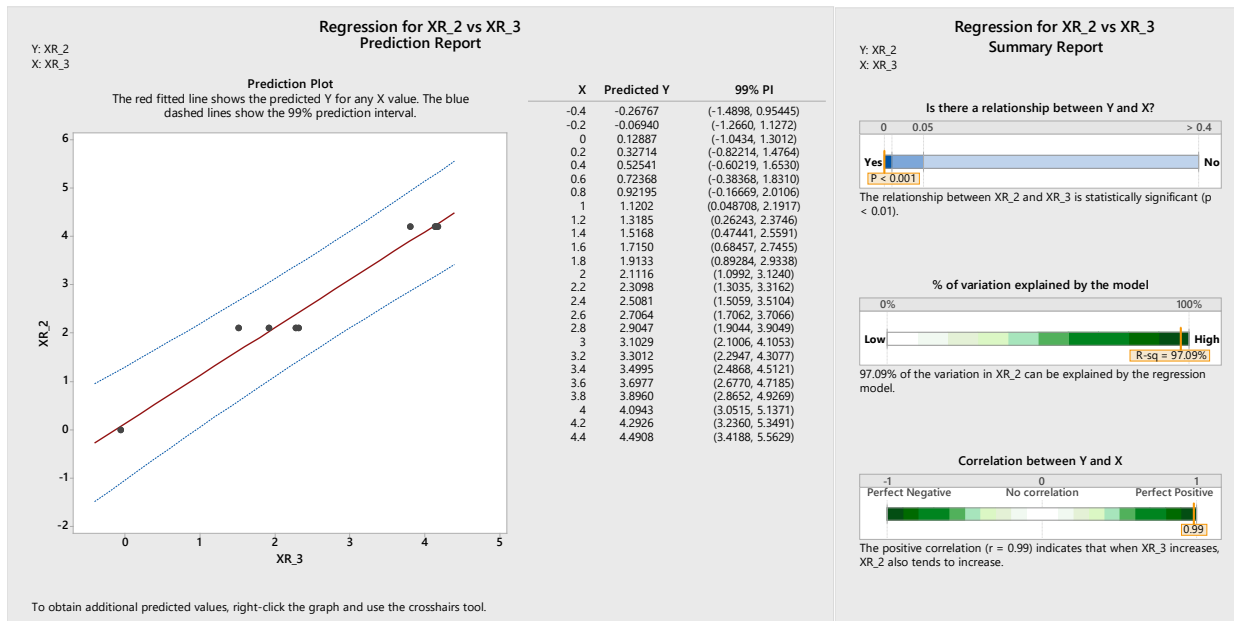


Figure III-25: X_R for Ductile Iron Regression Prediction Report

The XR model for ductile iron pipelines show a viable prediction platform as in Figure III-25. The model, marked as a red line, has a linear behavior with minimal deviations from the exact

experimental value. The model also presents a high statistical significance via the p-value which is less than 0.001. The r-squared value of the model is approximated at 97.09% with a correlation factor of 0.99, which is a considerably high value that attests for the statistical compatibility of the developed model. Since the model has a high value of r-squared and a p-value less than 0.0001, the next step is to conduct the F-test as in Table III-6. The model presents an F-value of 416.82, which attests to the statistical significance of the model because the value is substantially high. Furthermore, the used variables imply minimal to no impactful change in the p-value throughout the model development progression.

C. PVC Specific Regression Model

As in the ductile iron pipelines experiments, a PVC model is similarly developed using solely the data collected from the PVC experiments. The models are displayed in equations III-13 and III-14. X_L has an equation that is composed of three variables left MIE, right MIE and D, the distance between the two sensors.

$$X_L = -2.1364 + 3.569\frac{L}{T} + 2.591\frac{L}{R} - 0.18478D - 11.11\left(\frac{L}{T}\right)^2 - 0.2058\left(\frac{L}{R}\right)^2 \quad \text{(III-13)}$$

The predictive capabilities of equation III-13 can be summarized in Figure III-26 and the fifth row of Table III-6. The linear-shaped model shows that it is capable of predicting the leak location with minimum deviation and high accuracy. The statistical significance measured by the p-value is less than 0.001, which implies a high significance. The correlation is high for the model as r-squared is 98.96% and the correlation factor is at 0.99. Since all the preceding values prove the significance of the model, the next step is to check the F-value in Table III-6. The F-

value of the model at hand is equal to 637.22, which confirms the finding that the developed model is statistically significant.

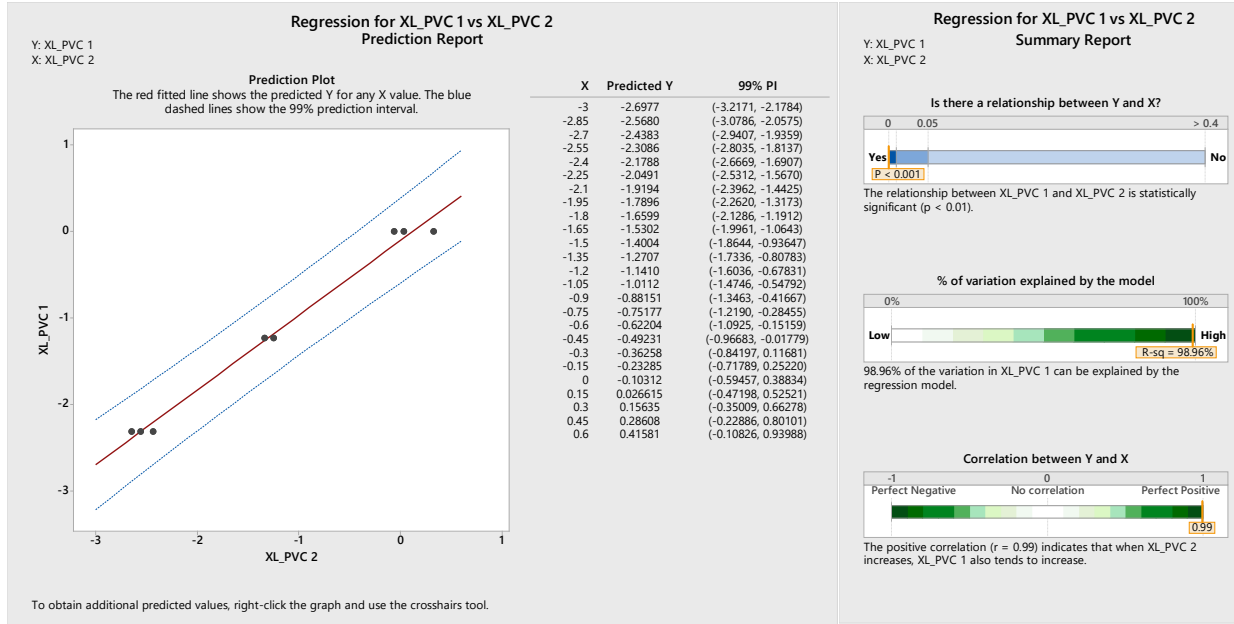


Figure III-26: X_L for PVC Regression Prediction Report

Additionally, none of the variables to develop the model in equation III-13 present a high p-value, completely negating the null hypothesis. In parallel, a model for detecting the X_R distance from the right sensor to the leak is developed. The model has a quadratic nature compared to all the previous models, as it is output by the Minitab software used for the development of all the regression models in this section. The X_R model in equation III-14 utilizes the MIE of the left and the right sensors along with the distance and their square variations.

$$X_R = 2.748 - 8.221 \frac{R}{T} - 0.151 \frac{R}{L} + 0.834D + 5.479 \left(\frac{R}{T} \right)^2 + 0.002344 \left(\frac{R}{L} \right)^2 + 0.03231D^2 + 0.01874 \frac{R}{L} D \quad (\text{III-14})$$

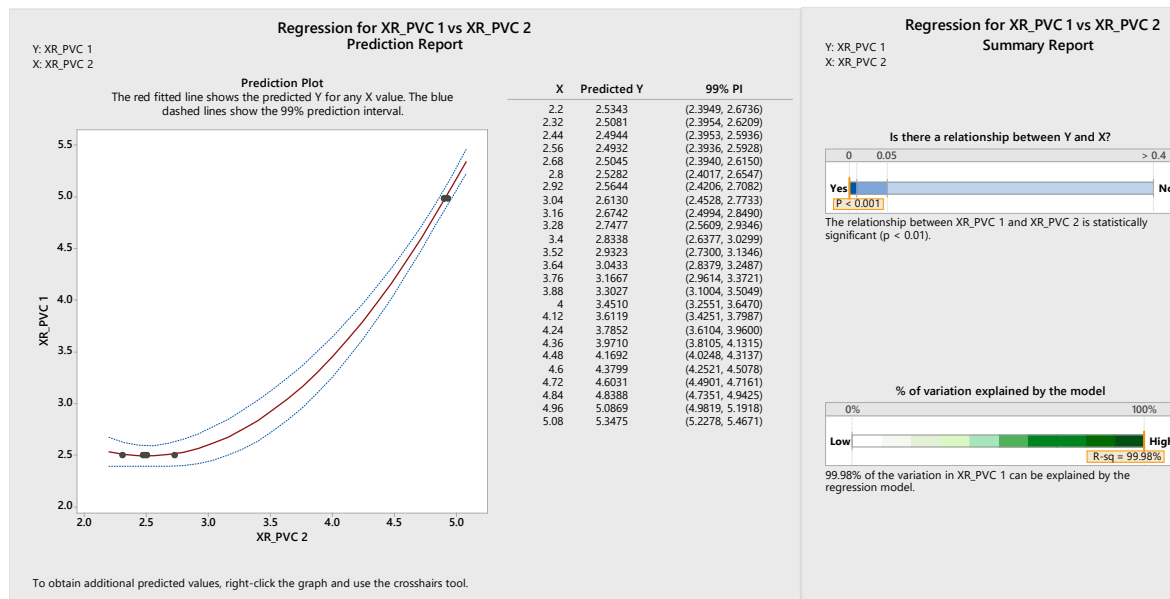


Figure III-27: X_R for PVC General Prediction Report

Figure III-27 presents the prediction report for the model of equation III-14. The figure confirms the quadratic nature of the equation on the contrary of its linear counterparts displayed in red within the graph of the figure. The model also shows a high statistical significance with a p-value less than 0.001 and value for r-squared that is equal to 99.98%. Accordingly, the F-value in Table III-6 is calculated to be 2360.5, proving the statistical significance of the developed model. Additionally, none of the variables had a p-value greater than 0.05 thus negating the null hypothesis.

D. Developed ANN Models

Comparatively, an artificial neural network model (ANN) is developed to test the extent of improvement of the regression models discussed in the previous section. ANN models are developed by the same variables selected by the best-subset analysis. Therefore, both techniques receive the same data input. Since ANN does not output an equation, the models are displayed in

Table III-4 regarding the mean square error (MSE) for both training and testing, as well as their relative r-squared in both training and testing. The models are developed using the Matlab neural networks package and the data is divided into 80% training and 20% validation. The summary of the stated statistical features of the ANN models can be found in Table III-4. All models performed relatively better than the regression analysis. The general model scores a mean square error of 0.06067 m² for the training of XL and a value of 0.04107 m² when validated. Those values represent an error of 20 centimeters, which is a very high accuracy value. Regarding R², the general ANN XL model scores a value of 96.92% in training and a value of 96.783% in testing to attest to the existence of a high statistical correlation. Similarly, for the general XR model, a low mean square error is scored in both training and testing with values of 0.05473 m² and 0.07265 m² respectively, thus averaging at around 25 centimeters. Consequently, the R² of the model in both training and testing is highly correlative, scoring a 99.5% value and a 98.71% value respectively.

Table III-4: Statistical Results of the Developed ANN Models

			XL	XR
General	MSE	Training	0.06067	0.05473
		Testing	0.04107	0.07265
	R ² (%)	Training	96.92	99.55
		Testing	96.783	98.71
PVC	MSE	Training	0.0344	0.0543
		Testing	0.0116	0.071
	R ² (%)	Training	98.64	99.1
		Testing	99.65	98.3
Ductile Iron	MSE	Training	0.0773	0.0627
		Testing	0.0897	0.09
	R ² (%)	Training	94.22	98.82
		Testing	96.91	97.75

In emulation of the performed regression analysis, two models are developed for PVC and ductile iron pipelines respectively. Besides, the ANN model for PVC pipelines performs better than its regression counterpart. The X_L model for PVC pipelines has a mean square error of 0.0344 m^2 and a testing MSE of 0.0116 m^2 . The value presents an average deviation of 19 centimeters, which is a high measure of accuracy for leak pinpointing. The X_L model of PVC shows a correlation percentage, R^2 , of 98.64% in training and 99.65% in testing. Similarly, an X_R model for PVC pipelines is developed. The model presents low mean square error values of 0.0543 m^2 and 0.071 m^2 in training and testing respectively, thus having a deviation of 26 centimeters in pinpointing the leak. The correlation of the model is also relatively high, having a percentage of 99.1% in training and a percentage of 98.3% in testing.

Finally, a ductile-iron model is developed. The ductile iron model performs less than its PVC counterpart with a higher mean square error on all ends. The ANN model for ductile iron performs much better than its regression counterpart. The X_L model has a mean square error of 0.0773 m^2 in training and 0.0897 m^2 in testing, thus having a deviation of 30 centimeters with a high correlation of 94.22% in training and 96.91% in testing. Consecutively, the ANN model for X_R pinpointing has a mean square error of 0.0627 m^2 in training and 0.09 m^2 in testing, thus having a deviation of 30 centimeters with a correlation of 98.82% in training and 97.75% in testing.

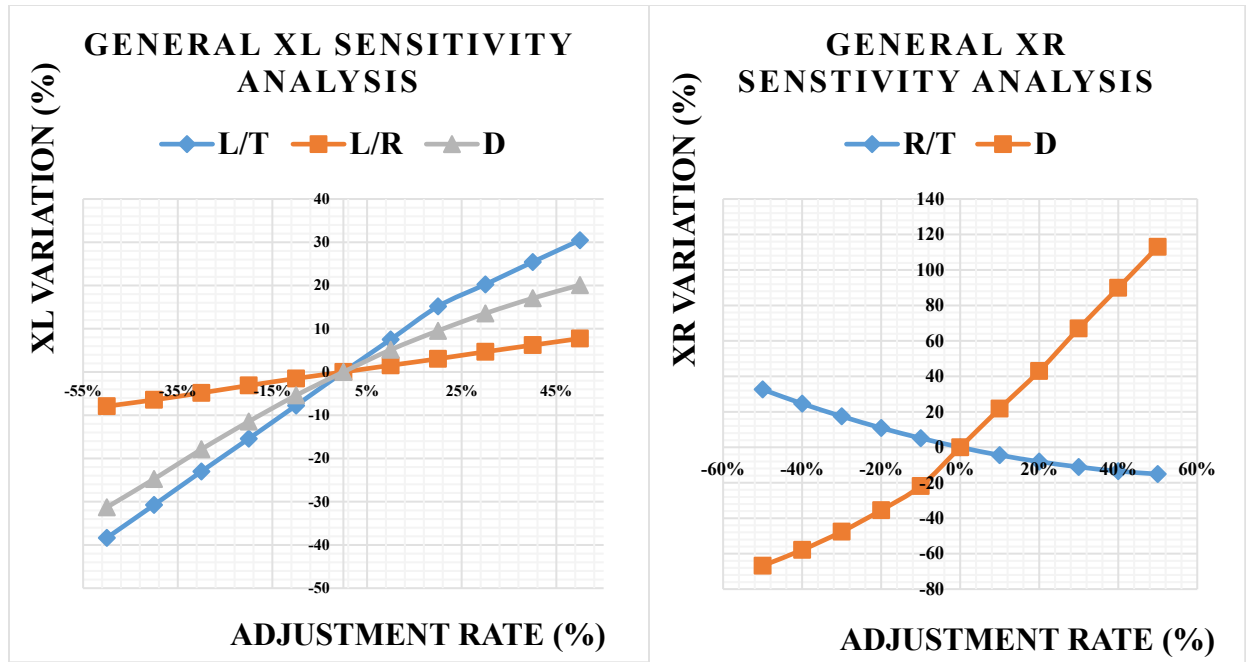


Figure III-28: Sensitivity Analysis of the General X_L and X_R Models

To delve into the understanding of the developed neural network models, a sensitivity analysis is performed for each developed model. The sensitivity analysis is conducted by varying the values of the parameters that comprise the model between -50% of decrease to +50% of increase with 10% increments. The model is run and the results are recorded to determine the level of change in mean square error. Figure III-28 displays the results of the sensitivity analysis conducted for the general X_L and X_R models. Figure III-28(a) shows the variation in the mean squared error as a function of the three variables used to develop the model, L/T, L/R and D. The general X_L model is most sensitive to the L/T parameter that represents the portion of the leftmost MIE to the total Mie in both sensors. This is because the slope for L/T is the steepest slope among the three parameters. Similarly, the general X_L model is the least sensitive to L/R due to a much subtle slope than the other two parameters. Figure III-28(b) presents the sensitivity analysis for the general X_R model against the two main parameters comprising it, which are R/T and D. The

two parameters have significantly high variations and impacts on the model, yet the distance between the two sensors, D , seems to have the highest impact on the model due to a steeper slope for its curve.

For specific models, a sensitivity analysis is conducted for the PVC and ductile-iron models. Figure III-29 displays the results of the conducted sensitivity analysis. Figure III-29(a) shows the results of the sensitivity analysis for the X_L model for ductile iron pipelines. The model is analyzed against its two parameters, L/T and D . The figure shows that both parameters have an equally substantial impact on the model because both parameters possess almost equal slopes for their sensitivity curves yet in opposite directions, making their impact inversely proportional. Therefore, the model is equally sensitive to its two parameters.

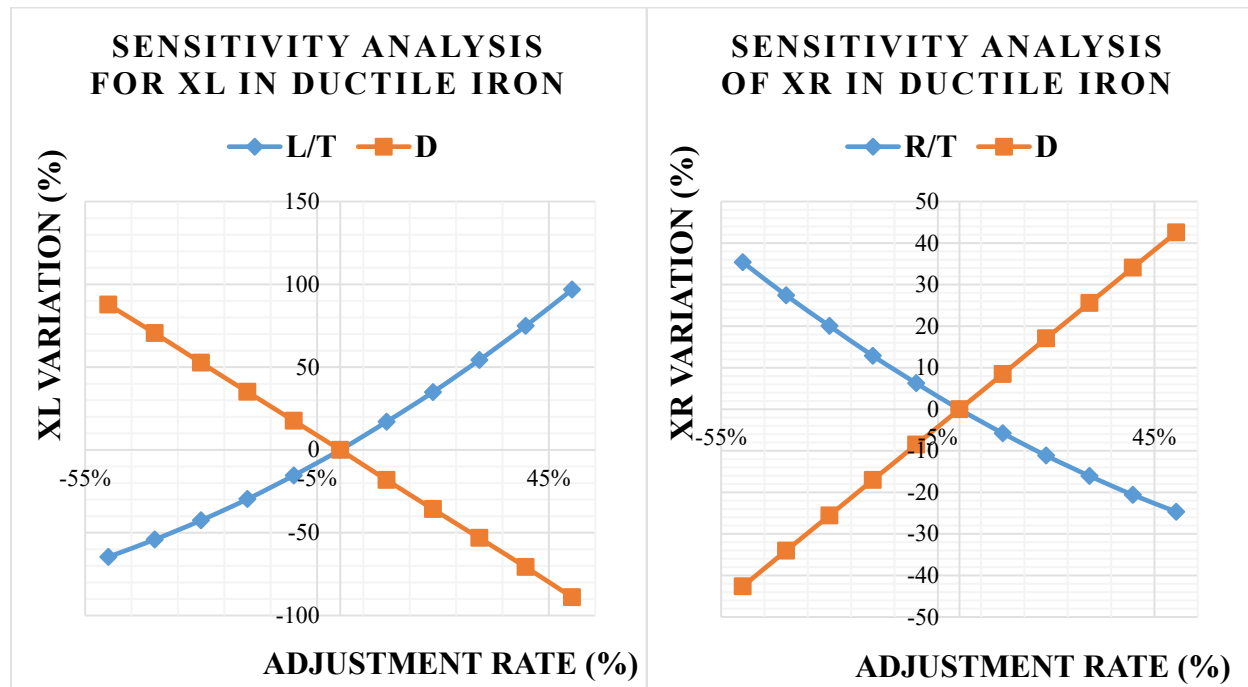


Figure III-29: Sensitivity Analysis of the X_L and X_R Models for Ductile Iron

Similarly, Figure III-29(b) shows the sensitivity graph for the X_R model developed using the ductile iron experimental data. As its X_L counterpart, the model shows the equal importance of its two parameters, R/T and D, with nearly equal slopes, yet with a slightly higher impact for distance D. Finally, for PVC, the results of the sensitivity analysis for the X_L and X_R models developed for PVC pipelines are summarized in Figure III-30. Both X_L and X_R for PVC are developed using three parameters. In Figure III-30(a), the sensitivity graph of the X_L model developed using the PVC pipelines is highlighted. The most critical parameter for X_L in PVC pipelines is L/T due to having the steepest slope amidst the parameters. On the other hand, D has the least steep slope between all three parameters, yet its impact is still considerably significant. As for X_R in PVC pipelines, Figure III-30(b) shows that the distance D is the most significant parameter holding the highest slope between all parameters. On the contrary, the impact of R/L on the overall model is next to zero because it has a very minimal variation regardless the direction in which it changes.

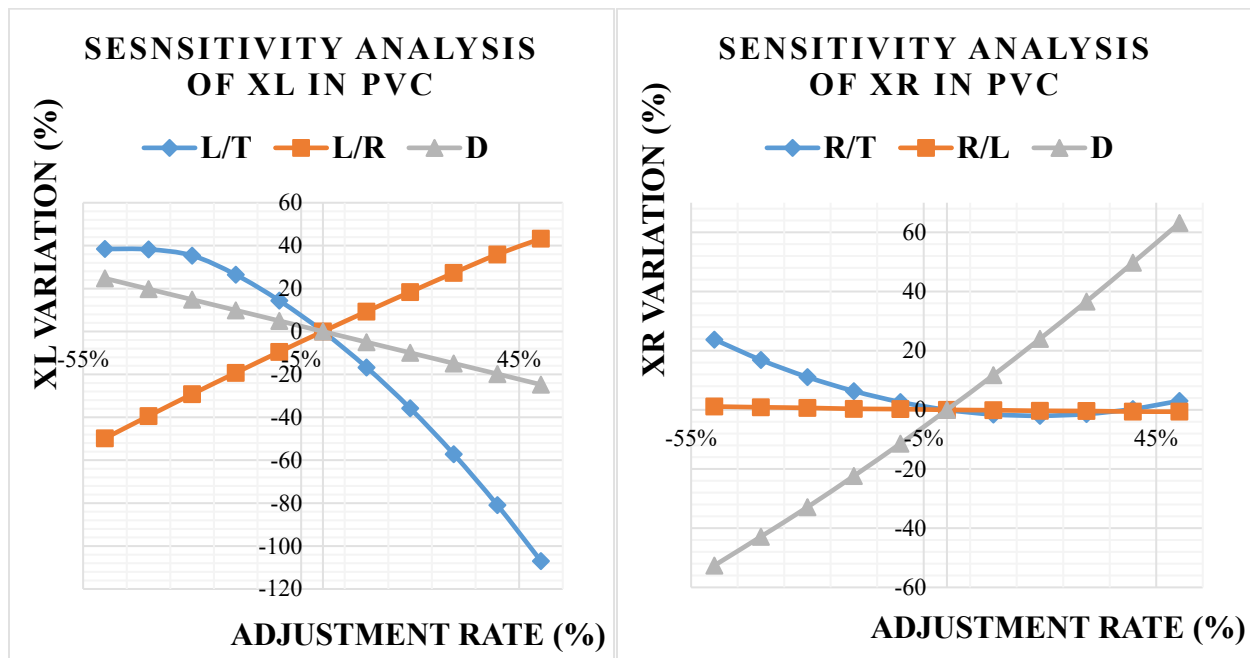


Figure III-30: Sensitivity Analysis of the X_L and X_R Models for PVC Pipelines

III.3.2.2 Statistical Analysis of Pinpointing Models

To select a model in regression analysis, multiple factors are taken into consideration. In this research, the utilized values are R-squared, R-squared predict, p-value, mallows C_p and mean square error (MSE). The selected models have the highest R-squared and R-squared predict, the closest mallow's C_p to the number of free variables and a low mean square error (El-abbasy et al. 2014). The model should also have a p-value less than 0.05 to negate the null hypothesis. Those values of the analysis are listed in Table III-5. The results in Table III-5 display that r-squared and r-squared-predict values are above 90%, indicating a high statistical relationship between the developed model and the derived output. The p-values for both X_L and X_R are both less than 0.05, negating the null hypothesis. Finally, Mallow's C_p values for the parameters under study are near the number of indicators or variables from which the equations are derived.

Table III-5: Statistical Analysis of the Developed Regression Models

	General		PVC		Steel	
	X_L	X_R	X_L	X_R	X_L	X_R
R²	94.8%	98.77%	98.96%	99.98%	93.99%	97.09%
R² Predict	91.87%	97.81%	98.75%	99.32%	88.68%	96.81%
P-value	<0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Mallows C_p	3.9	4.0	4.0	4.0	2.1	2.1
MSE	0.2444	0.2508	0.0829	0.0828	0.2535	0.2537

Table III-6 presents further insight into the statistical robustness of the proposed models in this section. The table contains a coding sequence that is followed by the number of the respective equation in this thesis. For each number, R^2 is restated and the p-value of the overall model is

identified. Having a high R^2 along with a p-value that negates the null hypothesis, i.e. less than 0.05, the models can be checked for their f-value. The overall f-value should accordingly have a high positive value to confirm the previous findings and highlight the statistical significance. Finally, the p-value of the coefficients through the step-wise regression are listed to state the step by step significance of each variable.

Table III-6: Advanced Statistical Analysis of the Developed Regression Models

Code	Model Equation Number	R^2 (%)	Model P-value	F-value	Coefficient P-value
1	III-9	94.83	< 0.001	187	<ul style="list-style-type: none"> • Constant = 0.000 • L/R = 0.000 • L/T = 0.000 • D = 0.014 • $D^2 = 0.048$ • $(L/T)*D = 0.000$
2	III-10	98.77	< 0.001	734.5	<ul style="list-style-type: none"> • Constant = 0.000 • R/T = 0.000 • $(R/T)^2 = 0.000$ • D = 0.000 • $D^2 = 0.020$ • $(R/T)*D = 0.000$
3	III-11	93.99	< 0.001	81.17	<ul style="list-style-type: none"> • Constant = 0.000 • L/T = 0.001 • D = 0.166 • $(L/T)^2 = 0.008$ • $D*(L/T) = 0.000$
4	III-12	97.09	< 0.001	416.82	<ul style="list-style-type: none"> • Constant = 0.000 • $(R/T)^2 = 0.009$ • $D*(R/T) = 0.000$ • D = 0.002 • R/T = 0.000
5	III-13	98.96	< 0.001	637.22	<ul style="list-style-type: none"> • Constant = 0.000 • L/T = 0.000 • L/R = 0.000 • $(L/T)^2 = 0.000$ • $(L/R)^2 = 0.000$

					<ul style="list-style-type: none"> • D = 0.000
6	III-14	99.48	< 0.001	2360.56	<ul style="list-style-type: none"> • R/T = 0.000 • R/L = 0.000 • D = 0.000 • (R/T)^2 = 0.000 • (R/L)^2 = 0.000 • RD/L = 0.010 • D^2 = 0.000

III.3.2.3 Validation and Testing of Pinpointing Model

To test the model, 20% of the experimental data was collected and not utilized in the development of the regression model. Through implementing the model, the values of X_L and X_R are calculated and the predicted values are compared to the original value to determine deviation DE_i . The deviation of each instance i is determined using Equation (III-15), where the deviation of each instance is determined by subtracting the exact location of the leak from the predicted location to get a numerical that describes how far is the prediction from reality and that is done using the equations mentioned above, such as (III-16) and (III-17) for X_L and X_R respectively.

$$DE_i = X_{APi} - X_{Ai} \quad (III-15)$$

Where:

DE_i = the deviation of the prediction from the original leak location of case i .

X_{APi} = the predicted value of X_L or X_R at case i .

X_{Ai} = the original real-life value of X_L or X_R at case i .

A = left (L) or right (R) indication of the value of X predicted or actual.

Furthermore, deviation values are collected and their standard deviation is determined and reported as σ_{error} . Another estimated value is the average deviation of the results of the testing set,

and that is \bar{D} , which is the summation of the deviation value of all tests over the whole number of values in the testing set. In order to determine the efficiency of the model in having the predicted value of the leak within 25 cm from the actual value, an equation is developed. This value is $A_{\pm 25 \text{ cm}}$. To calculate $A_{\pm 25 \text{ cm}}$, a variable A_i , a binary value of either one or zero, is determined. A_i is equal to 1 if the value of the deviation, D_i , is less than 25 cm. Otherwise, A_i is equal to zero. $A_{\pm 25 \text{ cm}}$ is the sum of all A_i over their total number n . The concepts of A_i and $A_{\pm 25 \text{ cm}}$ are illustrated in equations III-16 and III-17 respectively. The results and values are displayed in Table III-7. The output under consideration from the $A_{\pm 25 \text{ cm}}$ value is to study how many times is the developed model within the 25 cm range from the exact leak location.

$$A_i = \begin{cases} 1, & |D_i| \leq 25 \text{ cm} \\ 0, & |D_i| > 25 \text{ cm} \end{cases} \quad (\text{III-16})$$

$$A_{\pm 25 \text{ cm}} = \frac{\sum_{i=1}^n A_i}{n} \quad (\text{III-17})$$

Table III-7: Validation Results of Developed General Model.

	General		Ductile Iron		PVC	
	X_L	X_R	X_L	X_R	X_L	X_R
$A_{\pm 25 \text{ cm}}$	80%	85%	87.5%	87.5%	75%	100%
σ_{error}	18.74 cm	19.19 cm	19.44 cm	19.36 cm	19.88 cm	12.16 cm
\bar{D}	-0.8 cm	1.9 cm	-4.38 cm	-4.63 cm	-6.42 cm	-2.75 cm

Table III-7 shows that the standard deviation for the general X_L model is 18.74 centimeters and, for the general X_R model, it is 19.19 centimeters. Thus, the models has 68.2% of the results within the vicinity of 19 centimeters from the original leak location. Additionally, the average

deviation for X_L is -0.8 centimeters and, as for X_R , the average error or deviation is 1.9 cm for the original leak location. Additionally, Figure III-31 displays the distribution of the deviation and shows that the majority of the predicted values would lie between ± 25 centimeters from the original leak source with the existence of multiple outliers that can be attributed to experimental error.

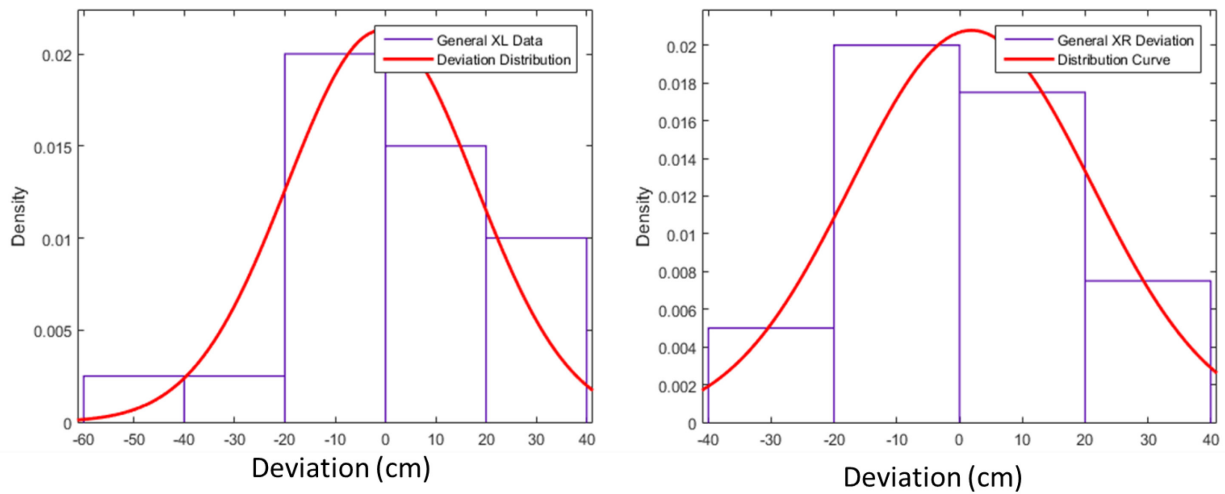


Figure III-31: Deviation Data Distribution for General X_L and X_R Equations

Similarly, Table III-7 shows that the regression models developed for ductile iron pipelines provide an accuracy of 87.5% for both X_L and X_R models. Additionally, the table shows that ductile iron accuracy is within ± 19.4 centimeters with an average distance deviation of around -4.5 centimeters. This is illustrated in Figure III-32 where the deviation of the model is generally within ± 20 cm. Additionally, the distributions of model deviation of X_L and X_R are very similar due to the closeness of the accuracies of both models as confirmed in Table III-7.

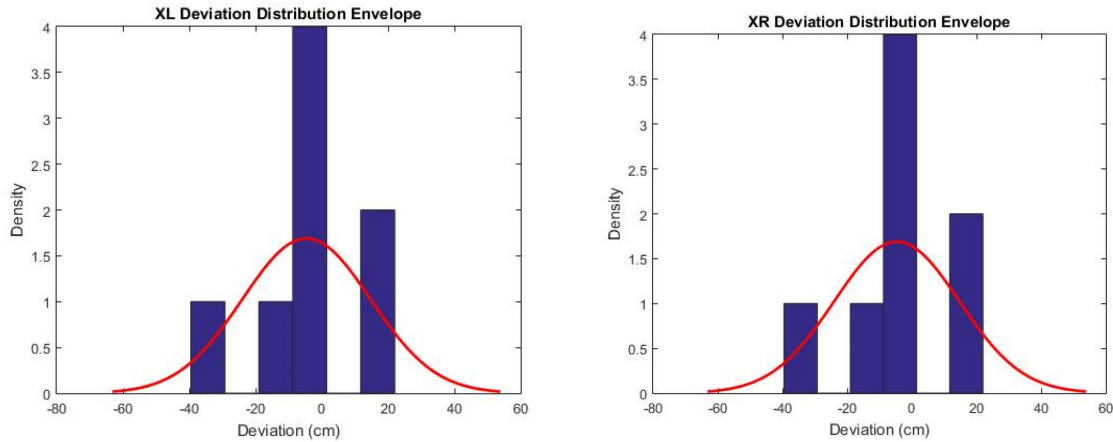


Figure III-32: Deviation Data Distribution for Ductile Iron X_L and X_R Equations

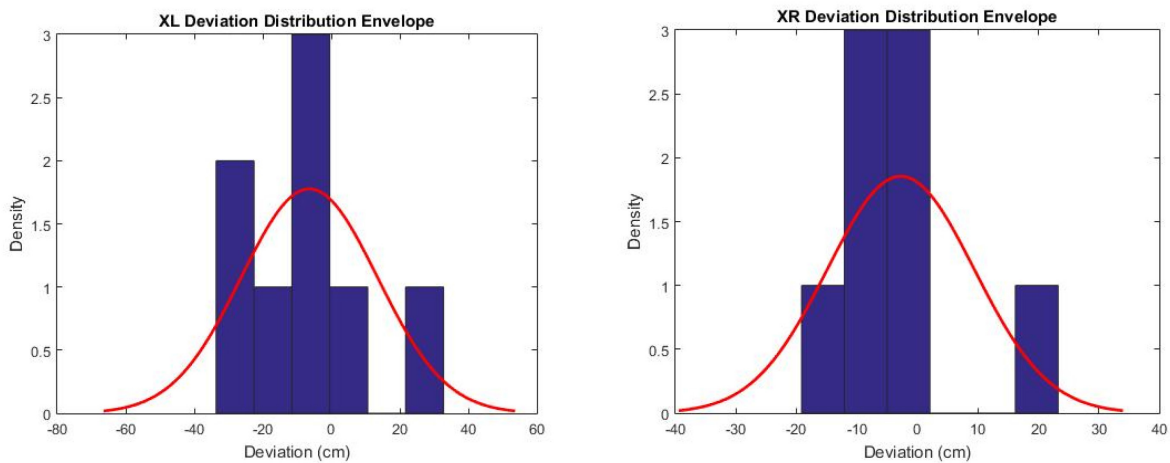


Figure III-33: Deviation Data Distribution for PVC X_L and X_R Equations

As for PVC pipelines, Table III-7 shows that X_L has an accuracy of 75% after testing, whereas X_R has an accuracy of 100% within a 25-centimeter range. The X_L model for PVC has a standard deviation of error equal to 19.88 centimeters, whereas X_R has a much minor deviation with a 12.16-centimeter standard deviation. The average deviation of the prediction of the X_L model for PVC is -6.42 centimeters and the average deviation of the prediction in the X_R model for PVC is estimated at -2.75 centimeters. Figure III-33 shows the results of the testing of the models. While the X_L model has a much wider range yet mostly within the limit of 20 centimeters, the X_R

model has a more precise and lesser range in prediction and produces more accurate results. The normal distribution curve is used as a prediction envelope for future uses of the model, to be a benchmark for predicting the accuracy of the model.

III.3.3 Vibration Signal Decay Model

The decay experiments are conducted to understand the nature of the vibration signal within the pipelines of experimentation. The aim is to see how the signal is lost within the system and in which direction it is lost. Table III-8 provides insight into the pattern where the signal is depreciating in terms of loss in monitoring index efficiency (MIE) or leak impact (LI). The study is conducted in two directions, to the left of the leak and the right of the leak. Leftwards, the signal travels against the flow of the pressurized water. Rightwards, the signal travels with the flow of water. Thus, we can also identify the existence of friction within pipelines.

Table III-8: Vibration Signal Decay Model Results

			MIE Decay		LI Decay	
			c _L (/m)	c _R (/m)	c _L (/m)	c _R (/m)
PVC	1-inch	μ	0.1	0.14	n/a	
		σ	0.043	0.024	n/a	
	2-inch	μ	2.14	0.4	2.62	0.52
		σ	0.5	0.18	0.4	0.18
Ductile Iron		μ	2.72	0.61	2.75	0.61
		σ	0.61	0.25	0.61	0.25

Table III-8 shows that for small one-inch PVC pipelines, the loss in MIE is insignificant, being around 0.12 MIE lost per meter. Although this value has an impact on the design and distribution of the monitoring network, the loss is minimal. Meanwhile, when the diameter of the PVC pipeline is doubled up to two inches, the losses against the flow of the water become highly more

critical and significant with a $c_L = 2.14$ MIE lost per meter in the left direction, whereas in the right direction, the loss is still minimal although more than that of a one-inch PVC pipeline with $c_R = 0.4$ MIE lost per meter. In PVC pipelines, leak impact estimates are higher than those of the monitoring index efficiency (MIE), which is returned to the tendency of the material to transfer vibration signals, making the baseline itself slightly more turbulent than that of ductile iron pipelines. Ductile iron pipelines have presented a higher tendency to dissipate vibration signals than PVC pipelines that tend to conserve the signal. Ductile iron has a c_L of 2.72 MIE per meter in MIE estimates in the left direction and a c_R of 0.61 MIE per meter. Since ductile iron is a more rigid material than PVC, no significant difference will be noticed if the decay value is measured for monitoring index efficiency (MIE) or leak impact (LI). Those values present a guideline for the distribution of the sensors within the network to be monitored. An example of the suggested distribution approach can be summarized in the following steps:

- **Step 1:** Select a leak threshold and an average leak size threshold.

In this example, the developed mid thresholds are selected such that a leak threshold is at $MIE = 1.052$ and a big leak threshold is at $MIE = 1.88$.

- **Step 2:** Find the value lost between the two thresholds.

In this example, the difference is $1.88 - 1.052 = 0.828$ MIE. The value represents how much of MIE needs to be lost before a sensor can no longer detect it.

- **Step 3:** Divide the value of the lost MIE over the signal decay in both flow directions.

The obtained value represents the required distance from the sensor to the average leak for it to be detected.

- a. **Case 1:** One-inch PVC pipelines.

- i- Leftward: $0.828 \text{ (MIE)} / 0.1 \text{ (MIE/m)} = 8.28 \text{ m.}$

ii- Rightward: $0.828 \text{ (MIE)} / 0.14 \text{ (MIE/m)} = 6 \text{ m}$.

b. Case 2: Two-inch PVC pipelines.

i- Leftward: $0.828 \text{ (MIE)} / 2.14 \text{ (MIE/m)} = 0.4 \text{ m} = 40 \text{ cm}$.

ii- Rightward: $0.828 \text{ (MIE)} / 0.4 \text{ (MIE/m)} = 2.07 \text{ m}$.

c. Case 3: Ductile iron pipelines.

i- Leftward: $0.828 \text{ (MIE)} / 2.72 \text{ (MIE/m)} = 0.3 \text{ m} = 30 \text{ cm}$.

ii- Rightward: $0.828 \text{ (MIE)} / 0.61 \text{ (MIE/m)} = 1.35 \text{ m}$.

- **Step 4**: Determine the desired length for distribution, based on the predetermined minimum and maximum. The determined leftward and rightward values represent the minimum and maximum required distances from a sensor in the given direction to the leak for a successful detection. To distribute the device efficiently on the levels of price and detection, it is recommended to choose a value near the average between the two determined values.

a. Case 1: One-inch PVC pipelines.

Selected distribution distance is 7 m. Thus, the distance between one sensor and the other is 14 meters for one-inch PVC pipelines.

b. Case 2: Two-inch PVC pipelines.

Selected distribution distance is 1.2 m. Thus, the distance between one sensor and the other can be 2.4 meters for two-inch PVC pipelines.

c. Case 3: Ductile iron pipelines.

Selected distribution distance is 0.85 m. Thus, the distance between one sensor and the other can be 1.7 meters for ductile iron pipelines.

CHAPTER IV: NOISE LOGGERS BASED MODELS

Similar to Chapter 3, Chapter 4 introduces a set of leak detection and pinpointing models in water mains using a different static technology which is noise loggers. This chapters aims to fulfill the second objective of this thesis that is to develop and validate technology-based leak detection and pinpointing models.

IV.1 Noise Loggers Leak Detection System

Acoustic technology is another leak detection technology for study and development in this research. The selected devices are noise loggers as a very abundant technology in the market, that offer a reasonable range of detection and pinpointing relative to their price. Loggers are currently part of a pilot project in the city of Montreal for the automation of leak detection wirelessly.

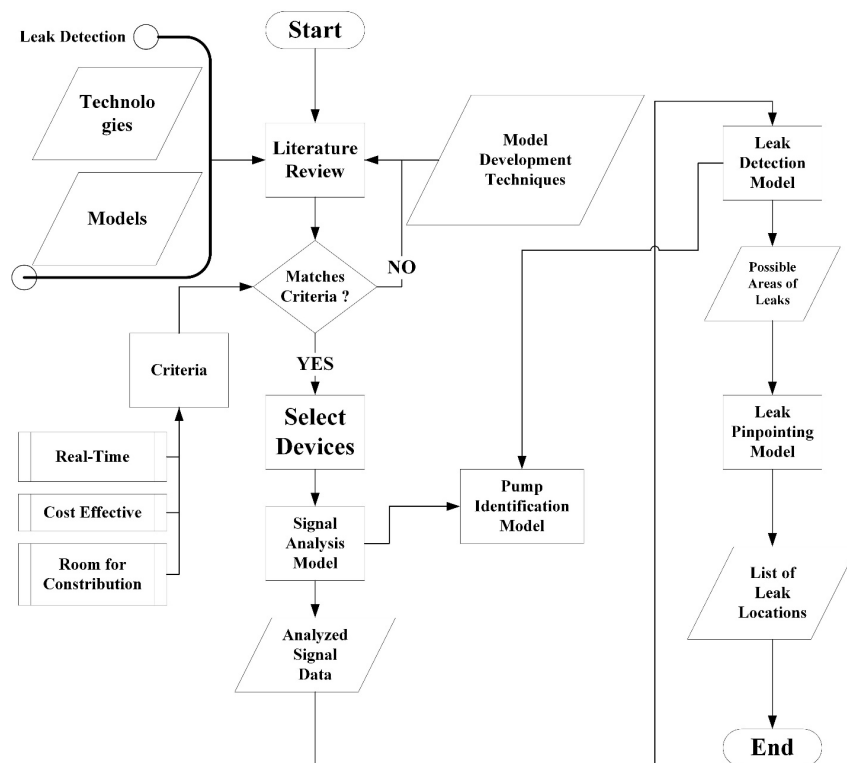


Figure IV-1: Acoustic Real-Time Monitoring System Model

The development of the acoustic real-time monitoring system has steps similar to those in the development of the accelerometer-based vibration real-time monitoring system. This is further clarified in Figure IV-1, where the system follows similar initial steps to those of the accelerometers. Yet, for noise loggers, there is neither a signal decay model nor a leak size estimation model to develop. Rather, a model is needed, that can identify acoustic signals that represent leaks and signals which represent pump sounds. The interference of signals has been highlighted in the city of Montreal as a critical issue in their pilot project, as causing a lot of false alarms, which in return cause an overdistribution of workforce and resources.

IV.1.1 Acoustic Signal Analysis Model

The conversion of signal sound files into comprehensible and measurable parameters is a significant task in this research since there is a need for values that can represent the cases under study. Figure IV-1 highlights the two forms of analyses conducted in this research: single signal analysis and dual signal analysis.

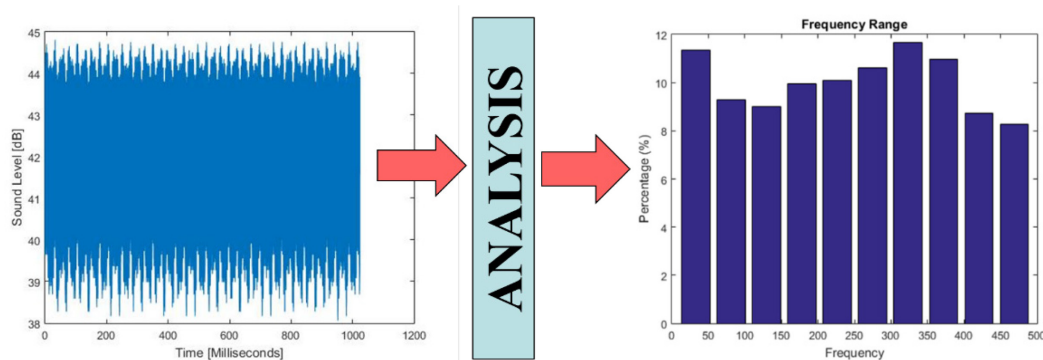


Figure IV-2: Frequency Distribution Analysis

The single signal analysis represents the full analysis of one single file where the required parameters are derived and developed. The first step for single signal analysis is Fast Fourier

Transform (FFT). FFT is conducted via a written Matlab code using the recorded sound file and converts the file into viewable and readable data represented in decibels, as displayed in the right image in Figure IV-2. Additionally, the FFT can provide the frequency distribution of the sound file, highlighting the repetitions and amplitude at each given frequency. The frequency distribution of a given sound file after analysis can be viewed in Figure IV-2, where the signal is converted to a percentage of the dominance of each frequency band starting with 0 Hz and ending with 500 Hz. The development of the acoustic signal analysis model is summarized in Figure IV-3. The first step is to determine whether the analysis is for a single file for the purpose of detection or more than one file for the purpose of pinpointing. If the analysis is for more than one file, then correlation analysis is performed to determine the relationships between the two sound files collected from the sensors surrounding the leak. However, a single signal analysis begins with conducting fast Fourier transform to determine the amplitude and the frequency curves of the signal.

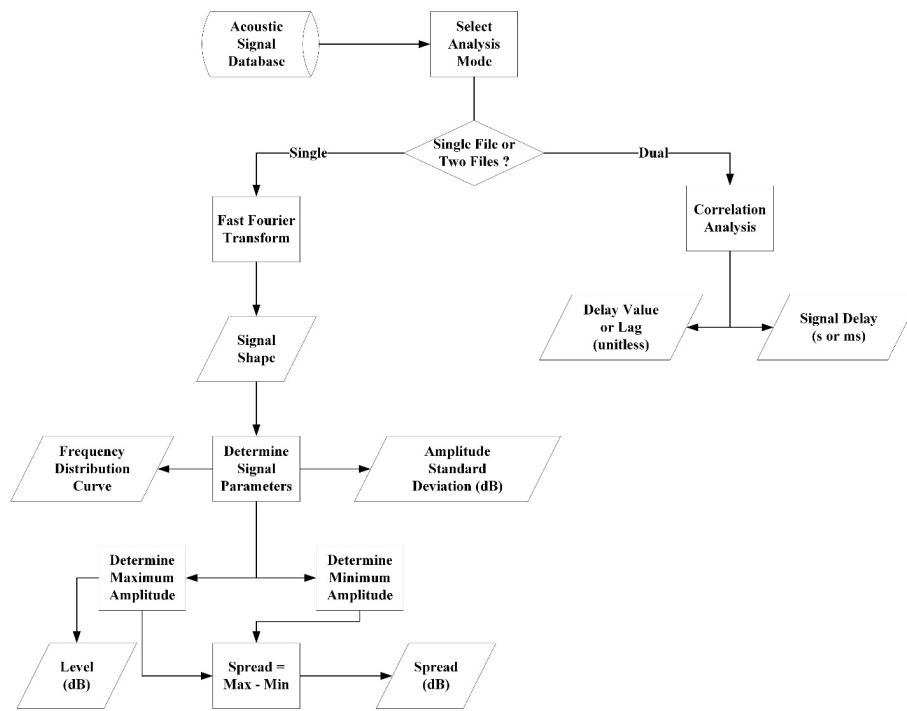


Figure IV-3: Acoustic Signal Analysis Model

The next step is to calculate the average amplitude as displayed in Equation **IV-1**. The value of the amplitude in decibels is calculated at each data point in the signal. The sum of those values is then divided by the total number of data points in the signal sequence.

$$\mu_{Signal} = \frac{\sum_{i=1}^N Amplitude_i}{N} \quad (\text{IV-1})$$

Where:

μ_{Signal} (dB) = the average sound amplitude of sound file under study.

$Amplitude_i$ (dB) = sound amplitude at instant i in the sound file under study.

N = the total number of data points within the sound file under study.

Following the calculation of the average sound amplitude, the next step will be to determine the standard deviation of the signal. Equation **IV-2** shows the calculation approach for σ_{Signal} by using the previously developed μ_{Signal} in Equation **IV-1**. The equation is the same as a basic standard deviation equation but rewritten to fit the particular problem at hand.

$$\sigma_{Signal} = \sqrt{\frac{1}{n} \sum_{i=1}^N (Amplitude_i - \mu_{Signal})^2} \quad (\text{IV-2})$$

Where:

σ_{Signal} (dB) = the standard deviation of the amplitudes within the sound file under study.

Following the development of the average signal amplitude and the standard deviation of the signal amplitude, the level of the signal and the spread are developed. The level of a signal represents the dominant or governing amplitude of the signal, i.e. the highest amplitude of the signal. Equation **IV-3** displays the calculation of a sound file's level through computing the sound amplitude in decibels at each instant i . The level of a signal is the peak of the signal. Thus, the equation determines the highest value in the sequence.

$$Level (dB) = \max(\{Signal(i) : i = 1, \dots, N\}) \quad (\text{IV-3})$$

Following the calculation of the level of the signal, another representation of the variation of the recorded sound is calculated. The parameter, in this case, would be the spread of the recording in the sound files. The spread represents the distance between the highest and lowest points in the signal sequence. Thus, the spread is determined by calculating the level or the maximum amplitude found in the recording as in equation IV-3. The level value determined has the minimum sound value subtracted from it. The spread represents the span of amplitudes within a signal.

$$Spread (dB) = \max(\{Signal(i) : i = 1, \dots, N\}) - \min(\{Signal(i) : i = 1, \dots, N\}) \quad (\text{IV-4})$$

After calculating the spread, the analysis moves from single sound file analysis to the analysis of two sound files representing the same phenomenon. In this case, the study aims at determining the level of similarity between the two signals by determining the correlation lag and also the time signal delay between the two signals. Regarding the correlation lag, the equations under study are discrete because the used loggers provide 16 seconds of recordings each time they record a sound and the values are consistent, so they do not increase or decrease. As a result, the discrete equation of correlation, Equation IV-5, has been used to determine the correlation lag value. The equation shows similarities with the convolution equation and convolution concepts. The aspired output of this equation is the displacement or lag (n) between the two signals. The

approach calculates the complex conjugate of function (f), (f^*), then moves ahead to determine the displacement (n) of function (f) against function (g).

$$(f * g) [n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m] g[m + n] \quad (\text{IV-5})$$

Where:

f = the initial sound function under study for the leak event.

g = the second correlation sound function under study for the leak event.

n = the lag between the two leak event functions.

Accordingly, after determining the lag values, another critical relational parameter is the time delay values. The standard cross-correlation equation for time delay, Equation IV-6, has been used to identify the discrepancy in time between the two leak event signals. The equation aims at determining the time difference between two signals at their point of similarity. The main unit for assessment is seconds. This value allows the user to realize which signal arrived before the other and with what time difference.

$$\tau_{\text{delay}} = \arg_t \max((f * g)(t)) \quad (\text{IV-6})$$

IV.1.2 Acoustic Leak and Pump Detection Model

To develop the approach, the first step was to collect multiple sets of data and to do experiments. The main topic was leak detection and thus two platforms were selected: acoustic signals and vibration signals. Acoustic signals were provided by the city of Montreal and were analyzed using Fourier transform and frequency spectrum in order to create detection baselines between

the leak state and the no leak state. One month of leak data was collected via acoustic noise loggers and was analyzed to form the acoustic leak database. When it comes to accelerometers, multiple experiments were performed to collect vibration signal data. Eight hours of signal data were collected and analyzed according to the Martini approach (Martini, Troncossi, and Rivola 2015). Using the analyzed and collected vibration signal data, the vibration signal database was constructed.

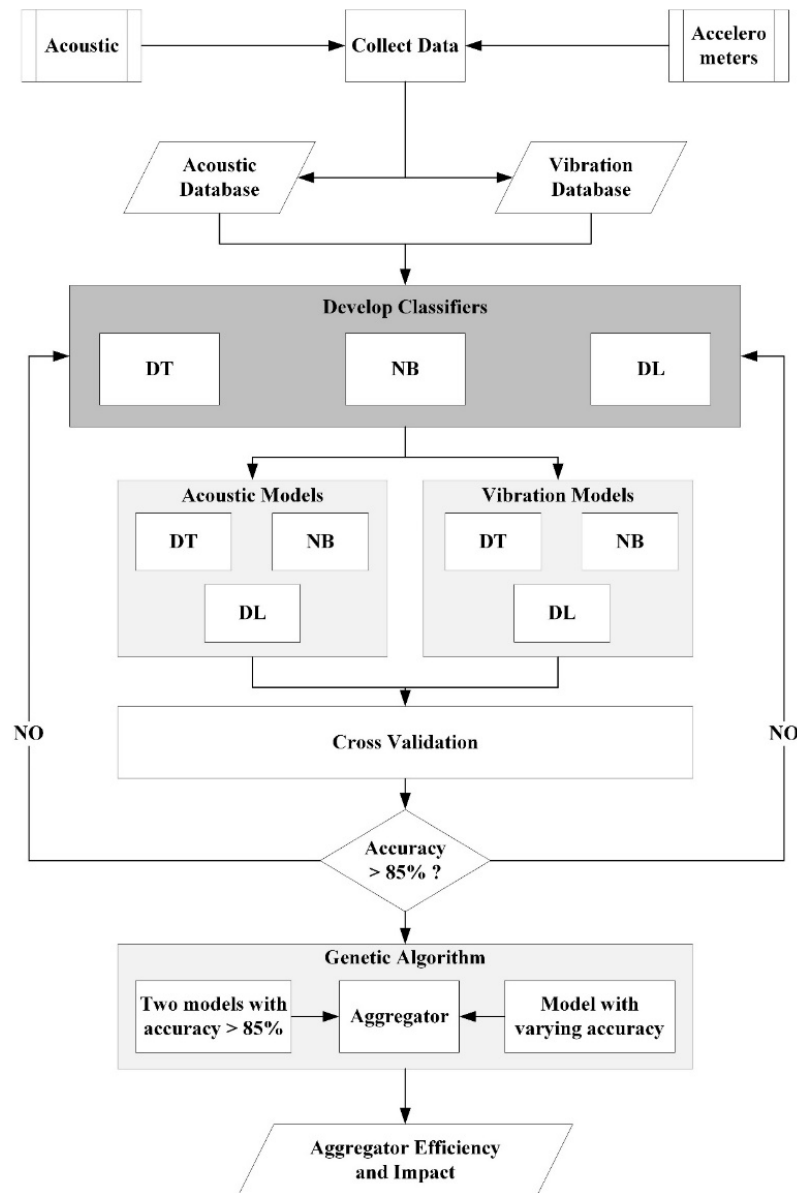


Figure IV-4: Identification by Aggregation Model

Figure IV-4 shows that a classifier was developed for each database using three different machine learning techniques with varying algorithms. The three techniques are namely Decision Tree, Naïve Bayes and 3- Deep Learning. Thus, acoustic noise loggers had three models with each technique and accelerometers. Each model was cross-validated with the test data set to check their accuracy. If the accuracy was above 85%, then the model would be accepted and used in further development. Otherwise, the model would be reassessed and redeveloped for further improvement.

The genetic algorithm operates on two instances to develop the envelope of improvement. Equation IV-7 shows the approach for determining the maximum possible accuracy using one varying model, where i represents the accuracy of the varying model and keeps changing between the values of 0 and 100 with an increment of 10% in each iteration. Then, the genetic algorithm is asked to modify the model such that when it is aggregated with the two available models, it has higher accuracy than in the initial model. The algorithm would run until all possible iterations are performed and the maximum possible accuracy – ***accuracyTotal*** – after aggregation would be determined.

For $i = 0:10:100$

Maximize: *accuracyTotal*

Subject to: *accuracyVariable = i* (IV-7)

End

Similarly, Equation IV-8 displays the approach for determining the minimum possible aggregated accuracy with the two good models. The accuracy of the varying model would be

varied between 0% and 100% with 10% increments, but the genetic algorithm aims to determine the worst possible scenario when aggregation is performed. Therefore, the output would be the minimum possible *accuracyTotal* of the aggregated model.

For i = 0:10:100

Minimize: *accuracyTotal*

Subject to: *accuracyVariable = i* **(IV-8)**

End

To assess the impact of the aggregation approach, two good models with accuracy over 85% were used against a varying model with no specific accuracy. The varying model would be varied using the genetic algorithm between 0% accuracy and 100% accuracy. The minimum and the maximum possible improvement would be recorded too, to form the envelope of the classification aggregator and therefore identify the advantages and disadvantages of aggregating multiple models.

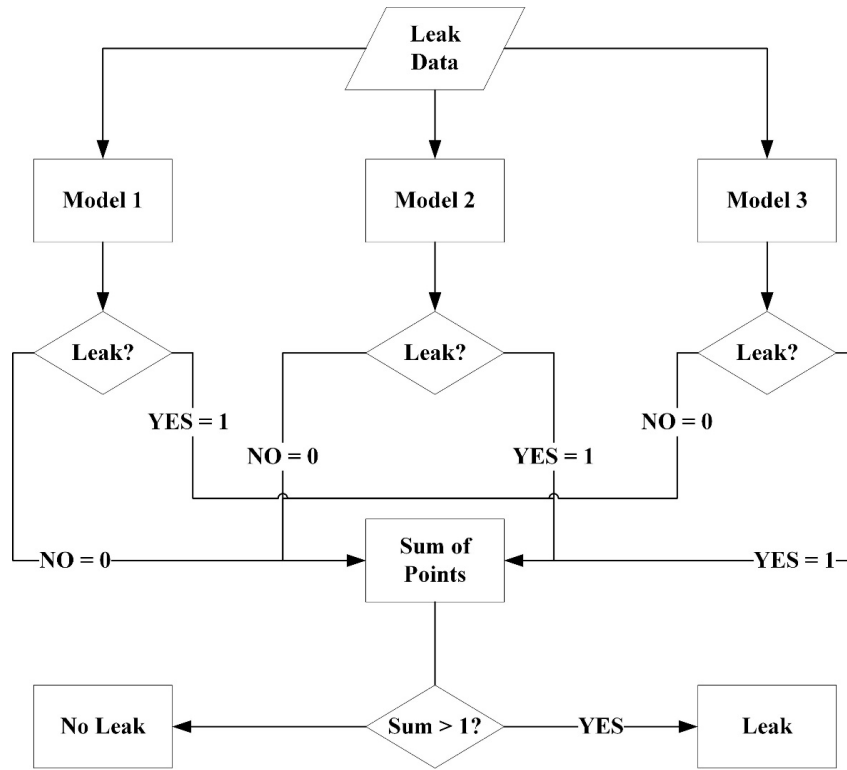


Figure IV-5: Aggregator Methodology

The architecture of the aggregator is simple and straightforward. The aggregator expects an odd number of classifying models, ideally 3 or 5. The aggregator will then monitor each classifier and record their response to a given data set. As shown in Figure IV-5, the aggregator was identifying between leak states and no-leak states. If the classifier identifies a leak, then one point would be recorded for the leak state. If the total recorded points for the leak state are two or three, then a leak is identified. Meanwhile, if the sum of points is 0 or 1, the data point would be considered to represent a no leak state.

IV.1.3 Acoustic Leak Pinpointing Model

Another critical aspect of the pilot project in the city of Montreal is the automated pinpointing of leaks using the sound recording of the noise loggers installed in the water network all over downtown Montreal. In Figure IV-6, the approach for tackling the pinpointing problem is

summarized. The initial step is to collect the analyzed signal data from two sensors prescribing one leak. The steps of signal analysis described in section IV.1.1 are conducted for dual signal analysis and for single signal analysis for each sound recording separately. Afterwards, best subset analysis for X_L and X_R is done for all the developed parameters based on the learning data set established from the recorded leaks and experiments. Based on the best subset analysis, the most suitable parameters for model development are selected and a regression analysis is conducted for X_L and X_R respectively. The developed models are then checked for accuracy, statistical soundness and the representation of the original dataset that was developed. In case best subset analysis approves more than one possible right approach for model development, multiple models will be developed. Each model that is deemed statistically sound and offers a good representation of the phenomenon would be added to a set of models for both X_L and X_R . The developed models would then be used as parameters to be regressed against the real data to develop a combining equation for all models and to present a final model for both X_L and X_R . When the newly developed final models for X_L and X_R and the total distance (D) between the two sensors under study are used, models for PX_L and PX_R are developed to help narrow the deviation gap and make the prediction of the leak location more precise. PX_L represents the percentage of X_L over D , i.e. the distance from the left sensor to the leak over the total distance between the two sensors. Similarly, PX_R represents the percentage of the distance from the right sensors over the total distance between the two sensors. The new models are then verified for accuracy and compatibility with the data that was collected for learning and validation. If the model is accurate and usable, it will be adopted and used.

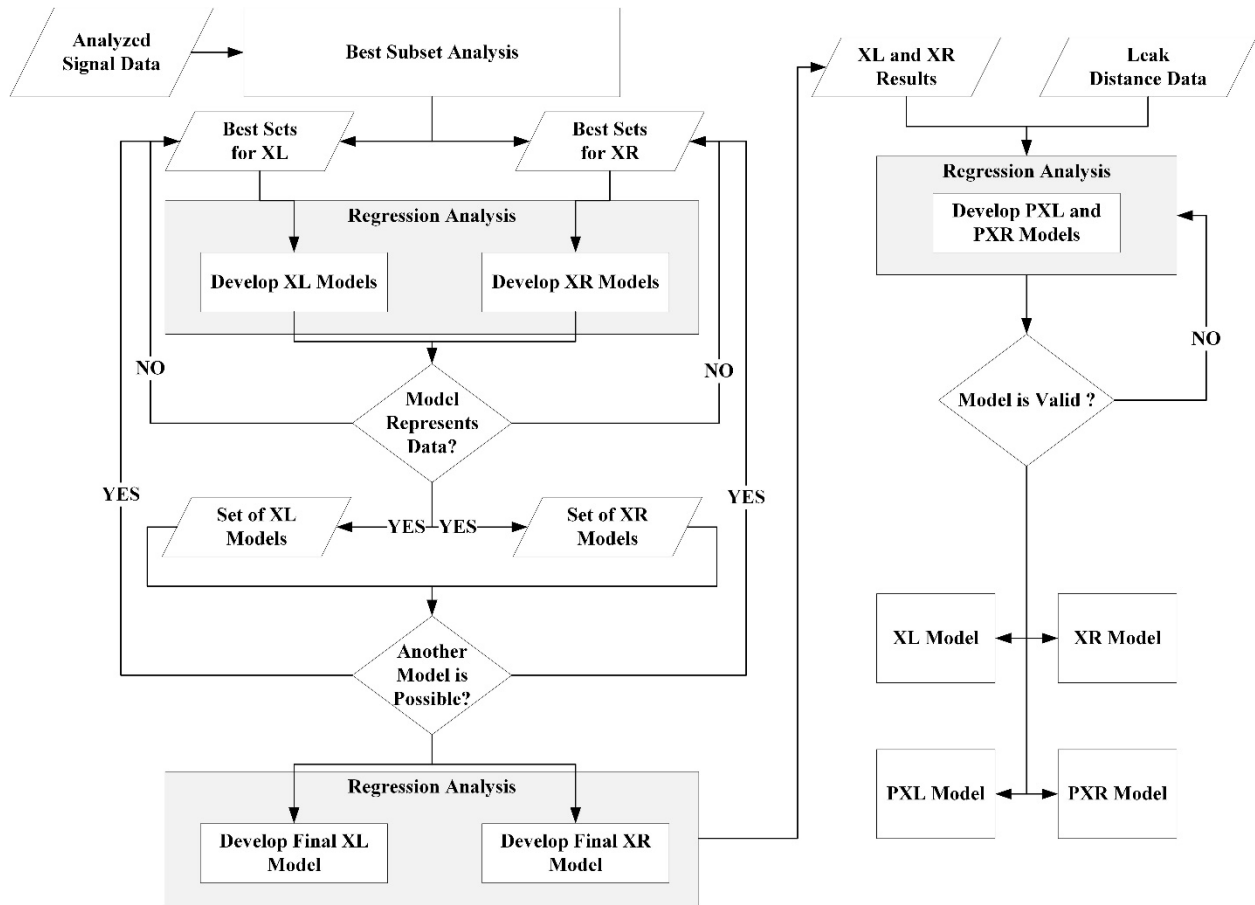


Figure IV-6: Acoustic Pinpointing Model Methodology

The proposed leak detection system differentiates from the leak detection system proposed by Hunaidi et al. in section II.1.4. by being real-time and therefore staying in the network all the time to detect leaks as soon as they happen, on the contrast of relying on leak suspicion to conduct a survey to verify the leak location as in the aforementioned model. Additionally, the model in section II.1.4 requires three main inputs to pinpoint leaks which are pipe type, pipe diameter, and the distance between the two sensors, whereas the model proposed in this section requires only the distance between the two sensors. The removal of pipe type and pipe diameter allows municipalities that lack proper records of their infrastructure to still benefit for early-on leak detection. Furthermore, LeakFinderRT system that is suggested in section II.1.4 relies on an integration between acoustic noise loggers and accelerometers, whereas the system proposed in

this section relies on noise loggers only, which helps in reducing the analytical and computational work done by the system. This can be highly beneficial in metropolitan areas where thousands of loggers are installed and multiple calculations are required at the same time.

IV.2 Data Collection for Loggers

The collection of data from acoustic loggers is done in coordination with the water services agency in the city of Montreal. The agency has placed a series of loggers covering downtown Montreal or the area known as Ville-Marie. The loggers are provided and installed by a company named Guttermann. The state of the loggers and the network can be viewed via an online system named Zonescan. Permitted by the agency, this research collects a substantial number of acoustic signals that cover multiple aspects of the acoustic leak detection spectrum. Additionally, the agency has checked to assess and verify the quality of their currently installed system.

IV.2.1 Zonescan System

The zonescan interface is an online system that connects all the installed acoustic loggers. Through transmitters and online servers, the data is collected and assessed to determine the existence of leaks, the status of the network and the status of the loggers and other devices forming the surveillance network. Figure IV-7 presents one of the interfaces of the zone scan system. This interface represents the status of the loggers regarding power and battery life. Similarly, through another interface, it is possible to detect leaks and collect sound files of the relative leaks. Additionally, through the regular checkups, it is possible to identify new sounds that create interference within the network, such as leak sounds and pump sounds. Another

interface is the leak detection interface, which presents correlated between multiple sensors to identify the existence and locations of leaks.

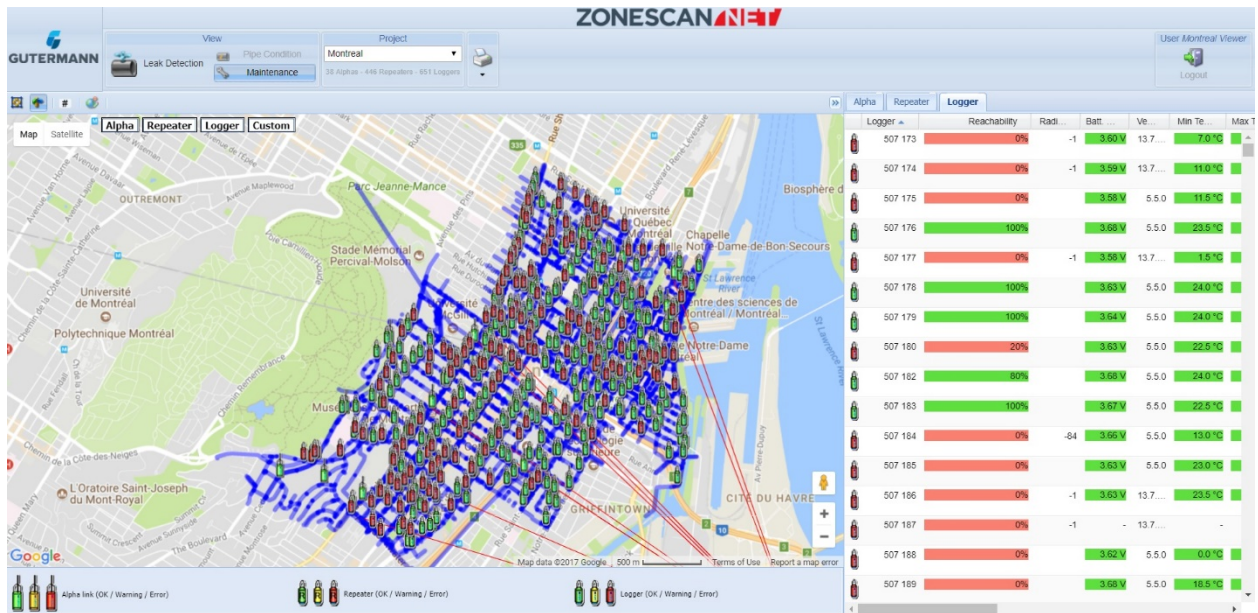


Figure IV-7: Zonescan Maintenance Interface

Figure IV-8 shows this interface by showing a detected leak in the vicinity of the Fine Art Museum of Montreal. To detect the leak, the figure shows that two loggers are used and their codes are presented as 507-241 and 507-244. The figure also shows the overall length of the pipeline between the two sensors as well as the predicted distance from each sensor towards the leak. Another presented measure is the distance from the center point between the two sensors towards the leak. The software indicates 100% certainty of the existence of this leak, but this remains to be further investigated as errors may arise from this system and that is one of the main reasons for the conduction of this research. The aim is to increase the reliability of this system through increasing its detection accuracy by eliminating false alarms and increasing its pinpointing accuracy as well.



Figure IV-8: Zonescan Leak Pinpointing Interface

To pinpoint a leak, the system uses the sound signals collected from the loggers on both ends of the suspected leak. Figure IV-9 shows that the system then moves towards performing a cross-correlation between the two signals for pinpointing purposes. As the figure shows, the software shows that the leak is closer to sensor 507-241 than to its counterpart sensor 507-244. Additionally, the software predicts that the distance from sensor 507-241 towards the leak location is 28.7 meters, whereas the distance from the second sensor, i.e. sensor 507-244, is estimated to be 56.5 meters. The software also indicates that the quality of their assessment is 100% and therefore the leak is highly likely to exist. Multiple situations may have occurred where pump sounds and other factors within the network create leak like conditions, but no leaks are to be found.

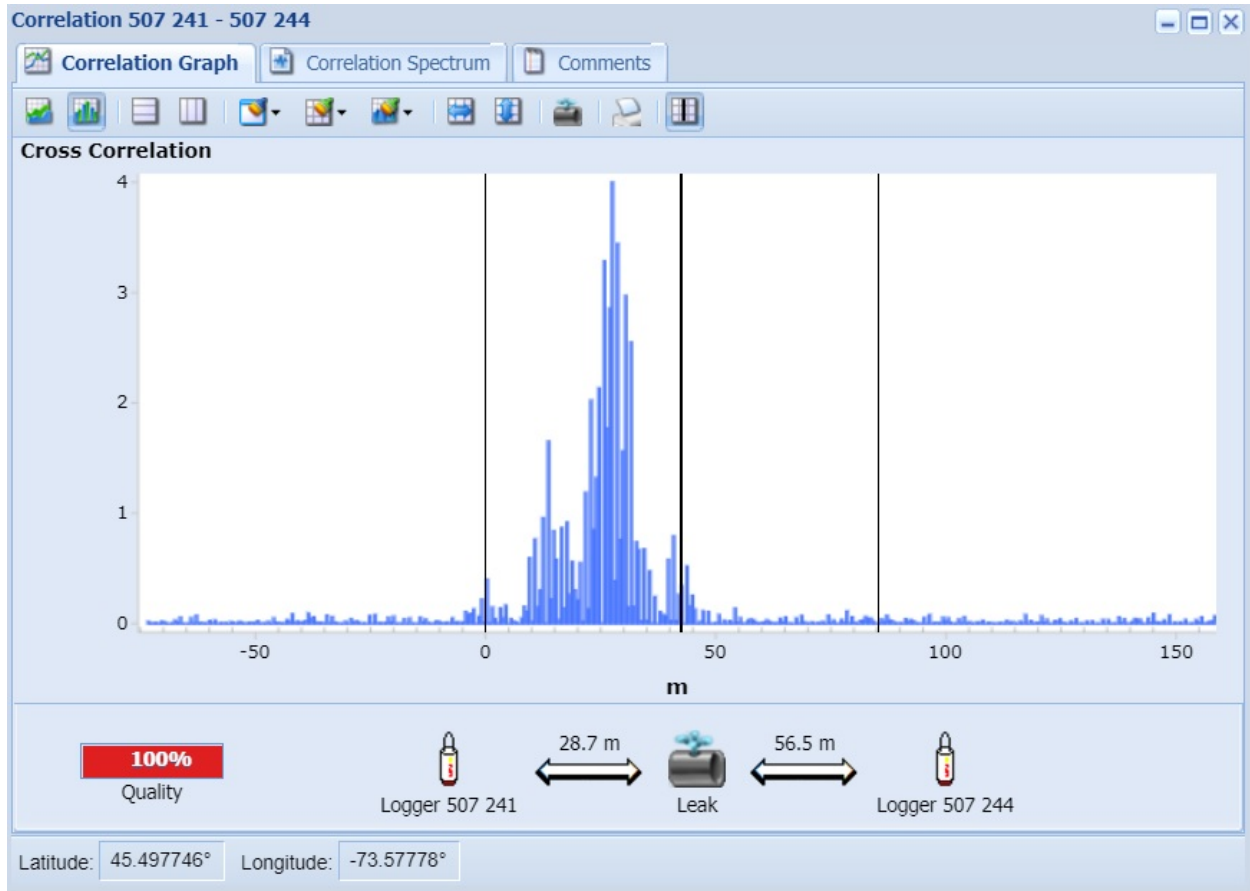


Figure IV-9: Zonescan Leak Cross Correlation Report

IV.2.2 Data Collection

The zonescan interface regularly listens to the underground network to detect sound anomalies within the network. The loggers start listening after 2 AM in the morning to minimize outside interference and noise. After listening for multiple hours, the loggers establish a sense of the ideal state and accordingly identify irregularities. In multiple cases, pump sounds are often assessed as leaks. The initial step is to collect sound files for various states. Therefore, the data is collected over the span of 30 days and various categories of sound files are collected, including leaks, no leaks, possible leaks, pump sounds and valve leaks. The possible leak option is created by the providing company. However, this research aims to eliminate the possible leak response

as it offers little to no information and adds confusion to the process. Additionally, pump sounds are distinguishable from leak sounds on the micro level because of the variation of the frequency distribution between the two classes of detection.

Over the course of 30 days, various types of sound signals were collected and analyzed, totaling 5167 sound files for different states of the leak detection system. Each sound file represents 16 seconds of recording and is comprised of 16384 data points forming the acoustic signal. Table IV-1 concludes the collection works over the course of 30 days, starting on April 1st, 2017 and ending on April the 30th of the same year. The collected files have 722 leak state sound files, 957 no-leak state sound files and 3488 sound files for possible leaks. The collected data is analyzed as in section IV.1.1 of the methodology and the data is collected and organized in an excel sheet for analysis.

Table IV-1: Summary of Collected Acoustic Data

Class	Leak	No Leak	Possible Leak	Total
Collected Signals	722	957	3488	5167

IV.3 Acoustic Leak Monitoring Implementation and Results

The main aspects of the leak detection approach in acoustic systems are the utilization of aggregation to minimize and even eliminate false alarms and the misclassification of the data received from the sensors placed all over the downtown of the city of Montréal. Consequently, in this section, the results of aggregation and its potential impact on any binary classification problem are discussed.

IV.3.1 Noise Logger Leak Identification Model

Multiple models are developed for the differentiation between leak signals and non-leak signals that are received from the noise loggers dispersed throughout the water network located in

downtown Montreal. Three different techniques are used due to their unique handling of data and processing of information.

The first model is developed using the decision tree algorithm. The decision tree differentiates between leak and no leak states by analyzing the frequency spectrum of each sound signal from both origins (leak and no-leak). The origin of the signals is confirmed by field investigation.

Figure IV-10: Decision Tree Model for Identifying Logger Leaks

and 200 Hz is greater than 18.395 million occurrences, then the algorithm may assume that the sound file is a leak. The decision tree is initiated by checking the frequency index for sound waves that are less than 50 Hz. Figure IV-11 presents the results of the cross-validation analysis that is conducted on the decision tree model. The model displayed an average accuracy of 85.89% with a possible deviation of 5.82% either positively or negatively.

accuracy: 85.89% +/- 5.82% (mikro: 85.87%)

	true Leak	true No Leak	class precision
pred. Leak	42	19	68.85%
pred. No Leak	21	201	90.54%
class recall	66.67%	91.36%	

Figure IV-11: Accuracy of Decision Tree Leak Identification Model for Noise Loggers

IV.3.1.2 Naïve Bayes Model

Using the same data set that was utilized for the development of the DT model, a Naïve Bayes model is developed. The Naïve Bayes model provided multiple distribution graphs to create statistical thresholds between leak and no-leak at each band of the frequency diagram. Figure IV-12 shows that the average cross-validated accuracy of the model is 83.35% with a possible deviation of 6.04%.

accuracy: 83.35% +/- 6.04% (mikro: 83.39%)

	true Leak	true No Leak	class precision
pred. Leak	54	38	58.70%
pred. No Leak	9	182	95.29%
class recall	85.71%	82.73%	

Figure IV-12: Accuracy of Naïve Bayes Leak Identification Model for Noise Loggers

IV.3.1.3 Deep Learning Model

The previous dataset was used a third time to develop a deep learning model to help differentiate between the leak and no-leak states. Figure IV-13 shows that the model had an average accuracy of 86.97% with a possible deviation of 6.38%.

accuracy: 86.97% +/- 6.38% (mikro: 86.93%)

	true Leak	true No Leak	class precision
pred. Leak	45	19	70.31%
pred. No Leak	18	201	91.78%
class recall	71.43%	91.36%	

Figure IV-13: Accuracy of Deep Learning Leak Identification Model for Noise Loggers

IV.3.1.4 Aggregation

The study is performed on three levels. The first level has two accurate models with one model of varying accuracy. The second level has one accurate model with two models of changing accuracy. Finally, the third level has three models of varying accuracies. These three stages aim to explore the possible aspects of binary classification improvement through aggregation.

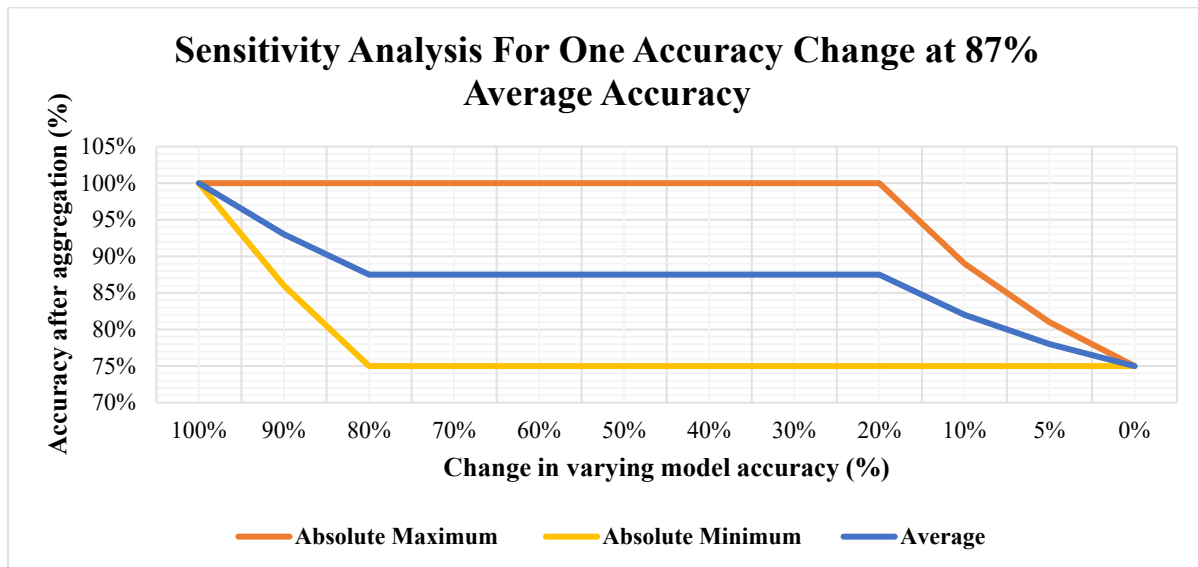


Figure IV-14: Accuracy Change Impact for One Model With Two Models at 87%

The first stage can be summarized in figures IV-14 and IV-15. Figure V-14 shows the possible envelope of aggregation for two models with an average accuracy of 87%. The figure shows that an overall accuracy of 100% is attainable with a model that is at least 20% accurate. This model must fit the missing gaps in the two other models, which could be challenging to provide. Therefore, to achieve a good aggregation impact, the lowest recommended accuracy for the third model should be at least 80%, beyond which the minimum possible outcome starts to decline, thus minimizing the potential for error. Equation IV-9 summarizes the findings in Figure IV-14. The equation calculates the expected aggregated accuracy, A_{agg} , of two models with an average accuracy of 87% when aggregated with a third model of accuracy A_{var} .

$$A_{agg}(2@87\%) = 0.7982 + 0.1536 * A_{var} \quad (IV-9)$$

Figure V-15 studies the amplification impact of a varying model when aggregated with two models of average accuracy of 87%. The figure shows that a model less than 13% of accuracy will have a negative impact on the overall model, whereas having a model with over 85% accuracy will be more influential and beneficial for the overall model when aggregated by providing an added accuracy of 10%, thus making the accuracy of the overall model 97%.

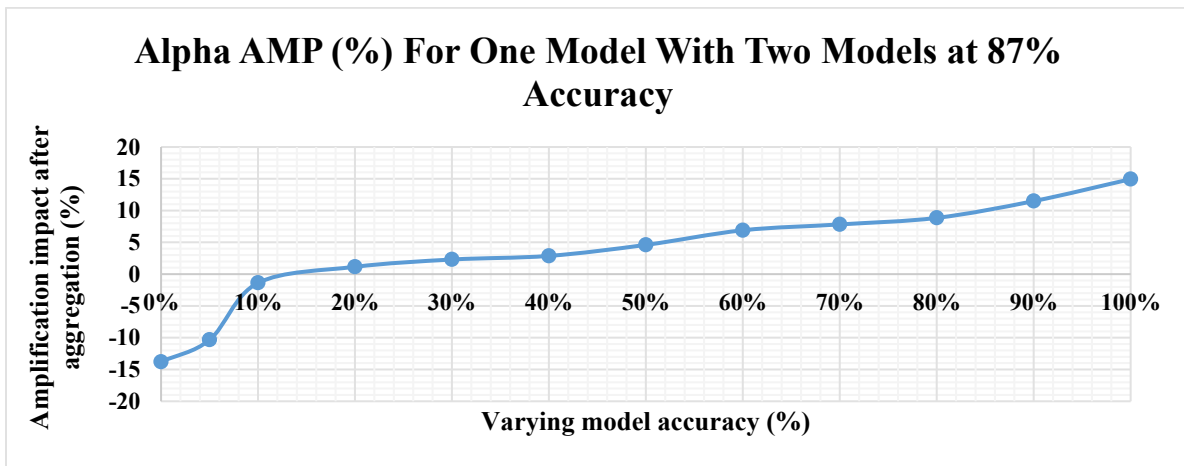


Figure IV-15: α_{AMP} for One Changing Model with Two Models at 87%

The second level includes studying the aspects and impacts of having one model with an accuracy of 89%, aggregated with two stochastic models with accuracies between 0% and 100%. In Figure IV-16, the envelope for aggregating one accurate model at 89% accuracy with two random models of the same accuracy is displayed. It is noticeable that the two random models have a varied potential, they can improve the model up to 100% accuracy with 60% accuracy each, but they can also destroy the model if they both carry 0% accuracy, for example. Therefore, it is highly recommended to utilize models with at least 60% accuracy to improve the accuracy of a model. Meanwhile, the more accurate a model, the more efficient the outcome of the aggregation. Equation IV-10 was developed from the average -yellow- line in Figure IV-16. The equation represents a mathematical model that can predict the aggregated average of a model with 89% accuracy with two models of equal accuracy A_{var} .

$$A_{agg}(1@89\%) = 0.02913 + 0.9443 * A_{var} \quad (IV-10)$$

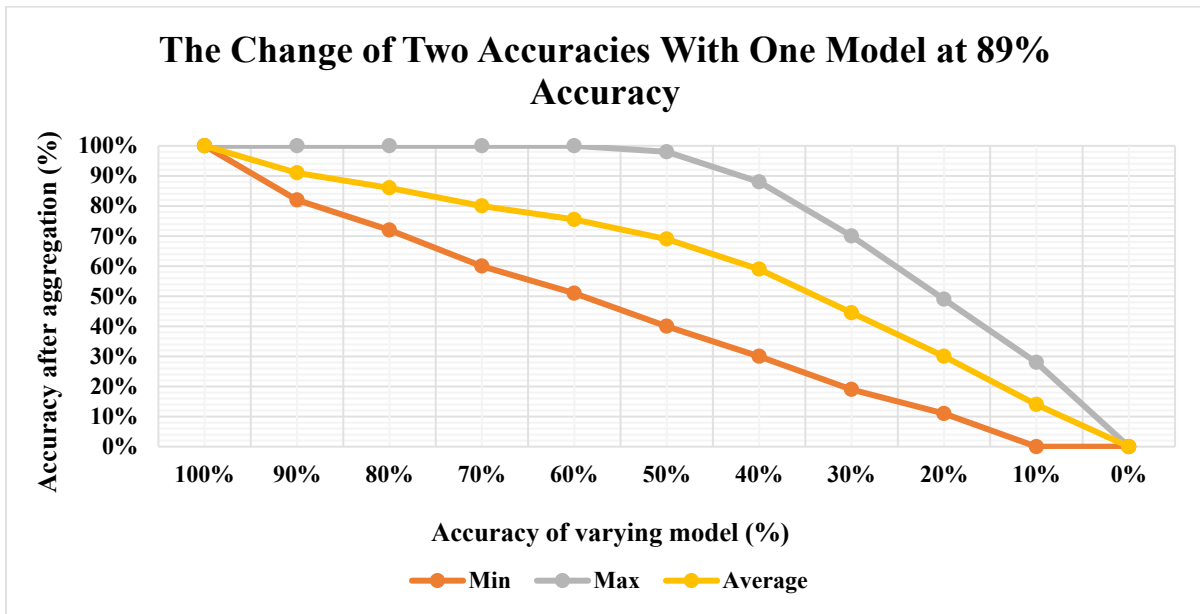


Figure IV-16: Impact Envelope for Two Models Aggregated with an 89% Model

Another studied aspect in this form of aggregation is α_{AMP} . Figure IV-17 shows that the model does not benefit much from aggregation until the two models are at least 88% accurate where the potential benefit above 88% value can reach help increase the accuracy of the aggregated output to 100%. Since the model is 89% of the time accurate, it is challenging to boost its accuracy with just any model and therefore the results of the figure present a valid point for this particular aggregation.

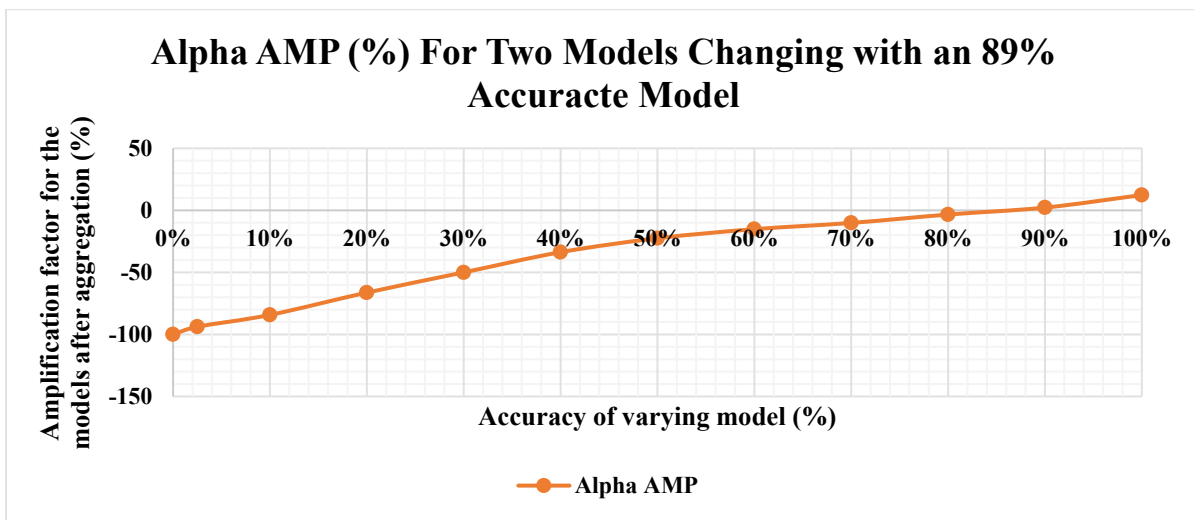


Figure IV-17: α_{AMP} for Two Models Aggregated With One Model at 89%

The last step in understanding the function of the aggregator approach is to utilize the three stochastic models and study the possible interactions between them. As Figure IV-18 shows, using three random models of the same accuracy opens multiple possibilities in the form of accuracy improvement. The figure also shows that 100% accurate models can be expected when models with accuracy equal to 70% or more are used. The figure also shows that, below 50% on average, the accuracy of the model worsens when aggregated and thus it is not beneficial to conduct aggregation below 50% accuracy. On the other hand, above 50% accuracy, models start

to improve by aggregation, and therefore it would be a recommended approach. The average expected accuracy, the yellow line, shown in the figure can be illustrated by equation IV-11. The equation shows that the predicted probability of an aggregated model using three random binary classification models of equal accuracies is a quadratic equation with A_{var} as its only variable.

$$A_{agg}(3 \text{ Random Models}) = 0.03066 + 0.2234 * A_{var} + 0.7562 * A_{var}^2 \quad (\text{IV-11})$$

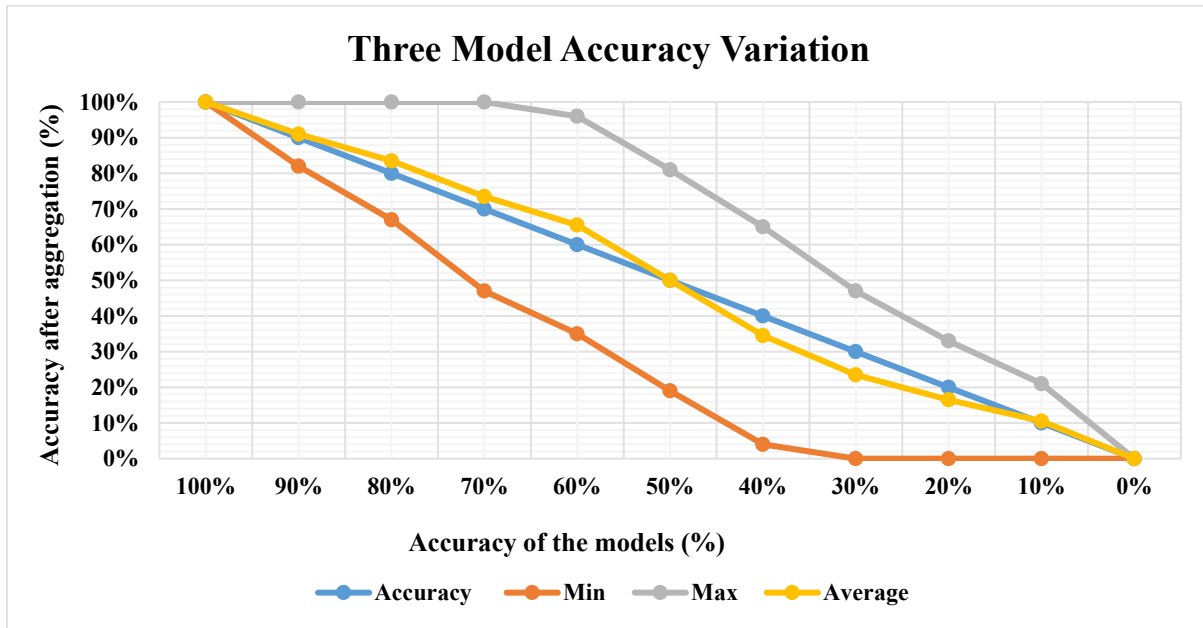


Figure IV-18: Aggregation Impact of Three Changing Models

Regarding amplification, the results of α_{AMP} are displayed in Figure IV-19. The figure confirms the deduction found in Figure IV-18, where no real benefit is found in aggregating models that are less than 50% accuracy. Additionally, it is noticed that models of 60% accuracy hold the highest rate of improvement with 9% added, if aggregated. On the other hand, models of 80% accuracy and above hold the highest value when improved by aggregation with the addition of an

extra 5% accuracy on average and 2% for 90% accurate models. The extent of the implementation of the aggregation model is tested in an acoustic leak detection and leak differentiation model. The model aims to differentiate between leak states and no leak states, but more importantly, to successfully identify leak sounds. Three binary classification models are developed using three techniques: Decision Tree (DT), Deep Learning (DL) and Naïve Bayes (NB). Those models are each presented with the same classification problem for an experimental data set of leak sounds and pump sounds. Table IV-2 shows the original accuracies of the models used in the development of the aggregation model. Decision Tree has an accuracy of 85.89%, Naïve Bayes has an accuracy of 83.35% and Deep Learning has an accuracy of 86.93%.

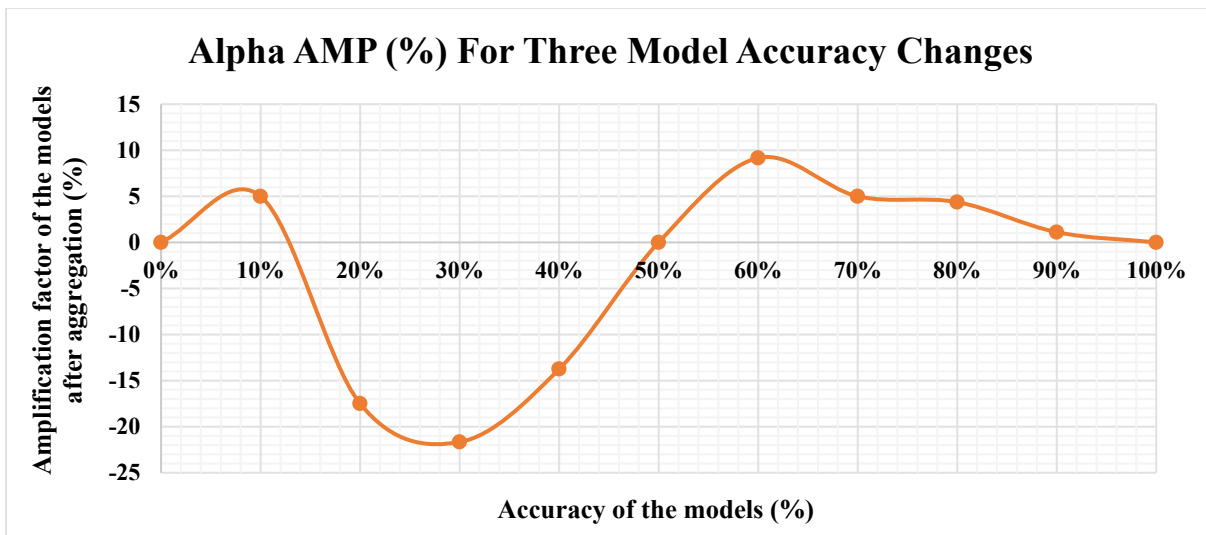


Figure IV-19: α_{AMP} of Three Models Changing Accuracies

Upon collecting the results of each model against the testing data set, they are aggregated to render the final decision and determine if the sound file represents a leak or not that the situation is typical within the network. The results of the aggregated values present 100% accuracy in detection and all the results are confirmed by the inspection that identifies the condition at the loggers and their respective pipelines.

Table IV-2: Aggregation Results of Leak Identification Model

Model	Decision Tree	Naïve Bayes	Deep Learning
Original Accuracy (%)	85.89	83.35	86.93
Aggregation Accuracy (%)	100		

IV.3.2 Acoustic Leak Pinpointing Model Implementation and Results

In this section, multiple steps are utilized for the development of a series of models for leak pinpointing in pressurized water mains using acoustic loggers. The initial step is to develop a model using best-subset analysis of the available data that is capable of pinpointing leaks. The second step is to use one model to detect its counterpart. Finally, a combination of the two models are utilized and X_L and X_R models are utilized to determine PX_L and PX_R , i.e. the percentages of the distances from the left and right sensors over the total distance between the two sensors.

The primary step is best-subset analysis. Figure IV-20 shows a sample of the best-subset analysis performed for X_L . The analysis consists of analyzing 12 variables derived from the analysis of correlated leak signals. The parameters describe the received signals and the relationship between those signals. The best subset analysis would then output a series of values that describe the statistical relationship between the desired output and the parameters at hand. The main aspects to study are Mallows Cp, R-squared and S. Mallows Cp represents the statistical relationship between the number of variables and the model. A good model is expected to have a Mallows Cp close to the number of variables plus one constant. The last model that is dubbed number 5 and utilizes all the available parameters has a Mallows Cp of 13, which is precisely equal to the number of variables 12 plus one constant. R-squared is the correlation factor that identifies how statistically significant the relationship between the variables and the output is. Therefore, the higher the r-squared, the better for the soundness of the model. Finally, S

represents the standard error for the prediction provided by the model in the same units as the desired output. For example, the final model with 12 variables shows a standard error of 1.9569 meters, implying that whenever this model makes a prediction, it is highly likely for the exact value to be within a range of ± 1.9569 meters.

Best Subsets Regression: XL versus D, FI 1, ...

Response is XL

						S i g n a l L a g S s V i i a g g l n n u S d a a e L S L p d B l l e p e r B d d (v r v e R e e l F F e e e a L i v v 0 I I l a l d e g d f h S S 8)											
Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	D	1	2	1	2	2	t	t	1	2)	
1	45.9	45.6	45.2	18370.2	19.340	X											
1	22.3	21.9	21.1	26482.1	23.185	X											
2	65.0	64.7	64.3	11812.4	15.592	X			X								
2	61.8	61.4	61.1	12925.5	16.299	X	X										
3	74.6	74.2	73.8	8528.3	13.321	X			X						X		
3	73.0	72.6	72.2	9082.9	13.739	X			X	X							
4	89.1	88.9	88.7	3553.8	8.7466	X	X	X	X								
4	87.2	86.9	86.5	4221.5	9.4961	X	X	X	X							X	
5	94.3	94.2	94.0	1765.1	6.3281	X	X	X	X							X	
5	94.2	94.1	94.0	1794.8	6.3763	X		X	X	X						X	
6	97.8	97.7	97.7	581.9	3.9660	X	X	X	X	X						X	
6	97.1	97.0	96.9	814.7	4.5331	X	X		X	X	X	X					
7	98.8	98.7	98.7	248.5	2.9722	X	X		X	X	X	X			X		
7	98.6	98.5	98.5	314.8	3.1959	X	X	X	X	X	X	X					
8	99.2	99.2	99.2	84.7	2.3265	X	X	X	X	X	X	X		X			6
8	99.2	99.2	99.1	94.4	2.3700	X	X	X	X	X	X	X			X		
9	99.5	99.4	99.4	8.2	1.9473	X	X	X	X	X	X	X		X	X		7
9	99.4	99.4	99.4	17.2	1.9951	X	X	X	X	X	X	X	X		X		
10	99.5	99.4	99.4	9.2	1.9471	X	X	X	X	X	X	X	X	X	X	X	1
10	99.5	99.4	99.4	9.8	1.9505	X	X	X	X	X	X	X	X	X	X	X	2
11	99.5	99.4	99.4	11.0	1.9515	X	X	X	X	X	X	X	X	X	X	X	3
11	99.5	99.4	99.4	11.0	1.9517	X	X	X	X	X	X	X	X	X	X	X	4
12	99.5	99.4	99.4	13.0	1.9569	X	X	X	X	X	X	X	X	X	X	X	5

Figure IV-20: Best Subset Analysis Report for X_L

IV.3.2.1 Initial Pinpointing Models

The first developed models for leak pinpointing rely on performing a best-subset analysis through all the available parameters for X_L and X_R . The models are represented in Equations IV-12 and IV-13. The primary variables under study are D , the distance between the two sensors that detected the leak, FI_n , the frequency index for each signal received from either the left or the right sensor, $Level_n$, the maximum sound level in decibels detected in the sound signal received from either the left or right sensors, $Spread_n$, the difference between the maximum heard sound in decibels and the least heard sound, and finally σ_{NS} , which represents the standard deviation of the signal decibels in either the left or right sensor.

$$X_L(1) = -7605 + 0.17604 * D - 15.098 * FI_L + 6.808 * FI_R + 189.57 * Level_L - 34.915 * Spread_L - 16.631 * Level_R + 11.446 * Spread_R + 14.89 * \sigma_{LS} \quad \textbf{(IV-12)}$$

The initial models present a good description of the leak pinpointing phenomenon and a good projection for the estimation of the location of future leaks. In Figure IV-21, the X_L model of equation IV-12 is displayed in red, surrounded by various data points used for testing and validation. The figure shows that the model has a limited number of outliers. On the statistical level, the model has a p-value that is less than 0.05, which attests to the statistical significance of the model in determining the location of the leak. Additionally, the figure shows that the r-squared of the model is estimated at 99.49%, proving that the equation is highly correlated with the reality of acoustic leak detection. All testing points within a 4-meter radius from the actual location of the leak and thus the accuracy of this model is $\pm 4m$. Table V-10 further proves the statistical significance of the developed model by providing an F-value of 3029.13, the high

value of the f-test when coupled with an r-squared that is close to 100% and a p-value that is less than 0.001. This negates the null hypothesis and confirms the robustness of the model.

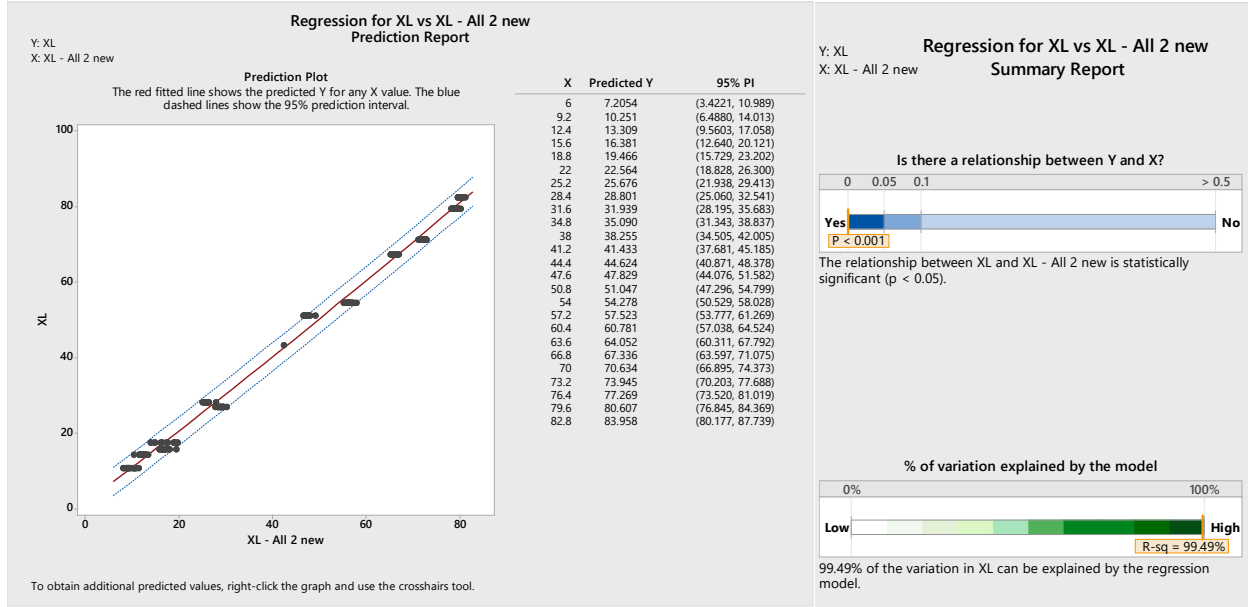


Figure IV-21: Acoustic $X_L(1)$ Regression Prediction Report

Furthermore, the table shows that the progression of the p-value has no fluctuations when all variables are utilized. Similarly, equation IV-13 describes the leak location from the right sensor to the expected leak location utilizing similar parameters to equation IV-12.

$$X_R(1) = -3887 + 90.57 * Level_R - 45.08 * Spread_R - 255 * \sigma_{RS} + 0.7614 * D - 5.96 * Spread_R * \sigma_{RS} + 7.96 * Level_R * \sigma_{RS} \quad (IV-13)$$

Figure IV-22 provides the analysis of the model. The figure shows the model in equation IV-13 as a red line surrounded by the testing envelope. The testing data shows that the model is usually within 19 meters from the exact leak location, thus having an accuracy of ± 19 meters. The model

presents a p-value that is less than 0.05 and therefore the model is statistically significant and capable of predicting the location of the leak in the right direction. Moreover, the model has an r-squared value of 91.69%, which shows a significant correlation between the model and the location of the leak. Finally, the model has a correlation factor of 0.96 between the predicted value and the exact value of the leak location. To complement the findings of the model analysis, the f-value of the model is calculated and listed in Table IV-3 to be 342.47, which is a value greater than zero and proves the significance and quality of the developed model. Additionally, throughout development, the utilized variables do not create any increase in the p-value of the model.

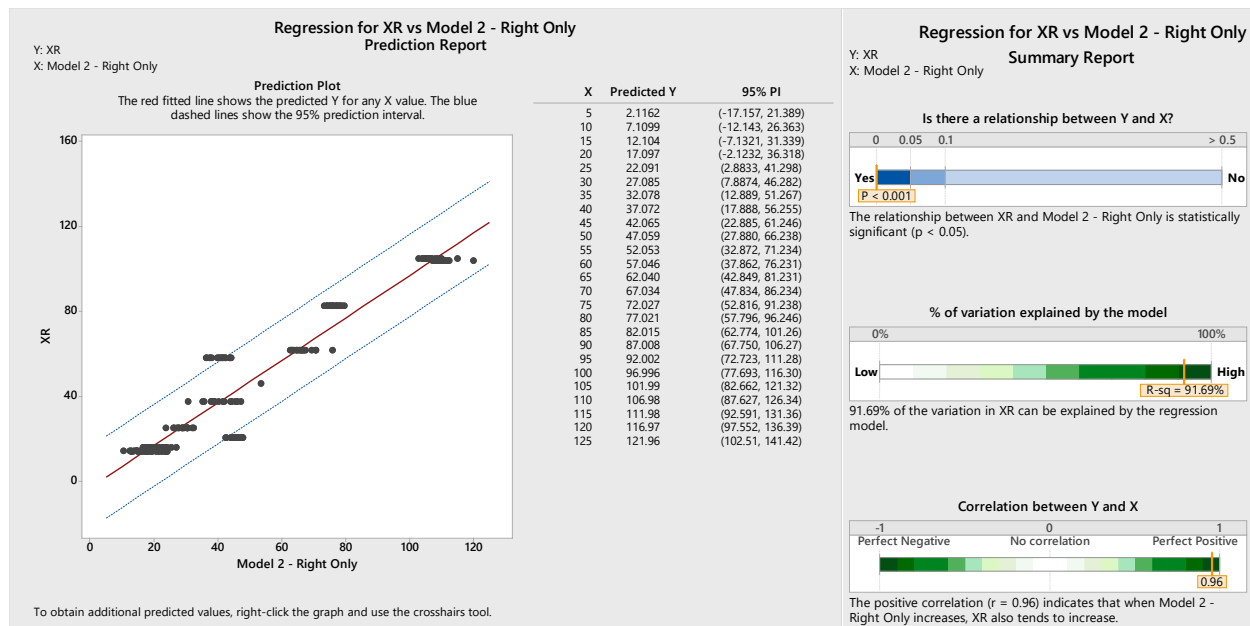


Figure IV-22: Acoustic $X_R(1)$ Regression Prediction Report

IV.3.2.2 Nested Pinpointing Models

The concept of nested models suggests using the results of the opposing initial model to deduce the complementary value for leak pinpointing. For example, the value of the original $X_R(1)$ is used in the regression of the $X_L(2)$. As for $X_L(2)$, the model is illustrated by equation IV-14. Through the regression of this model, the Minitab software repeatedly eliminates $X_R(1)$ and labels it as insignificant. Yet, due to the addition of the lag value between two signals, a new model is derived regardless of the contribution of the initial model.

$$\begin{aligned} X_L(2) = & 34.05 + 0.1276 * D - 15.06 * FI_L - 20.95 * FI_R + 53.82 * Level_L - 19.89 * \\ & Spread_L - 52.16 * Level_R + 35.07 * Spread_R + 39.74 * \sigma_{LS} - 29.22 * \sigma_{RS} - 53.35 * \\ & LagValue \end{aligned} \quad \textbf{(IV-14)}$$

The results and accuracy of the model is summarized in Figure IV-23, where the model is displayed in red surrounded by the testing data points for prediction. The model has a few outliers that impact the overall accuracy. The model also has a p-value less than 0.001 to validate the statistical soundness of the model. The r-squared of the model is assessed to be 97.08%, presenting a strong relationship between the variables and the desired output. Finally, the correlation coefficient has a value of 0.99 to display the high quality of the model in representing the phenomenon under study. The f-value in Table IV-3 confirms the findings of the r-squared and the p-value. The f-value of this model is estimated to be 2788.32, which is a value greater than zero. Additionally, the variables utilized in this model have minimal to no impact on increasing the p-value of the model.

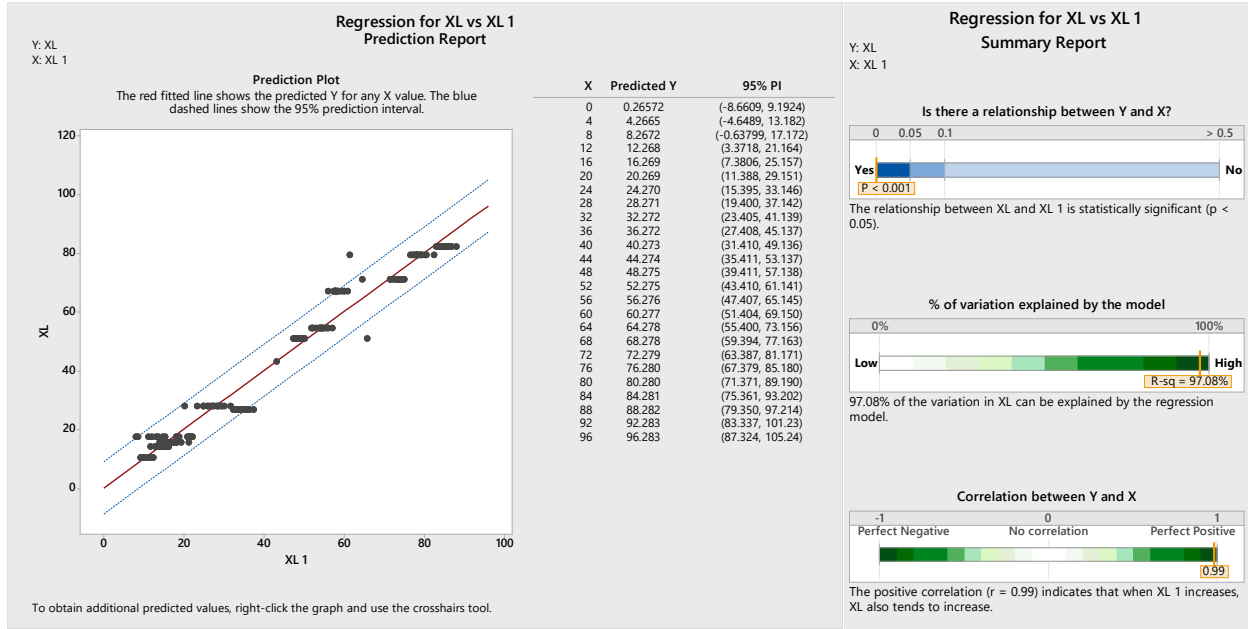


Figure IV-23: Acoustic $X_L(2)$ Regression Prediction Report

Similarly, a model for X_R is developed, named $X_R(2)$ and presented by equation IV-15. On the contrary of $X_L(2)$, $X_R(2)$ utilizes the value of X_L presented by the $X_L(1)$ model to determine X_R . Therefore, this model is a nested model having to use two regression models to achieve its final value.

$$\begin{aligned}
 X_R(2) = & -252.25 + 118.01 * FI_L + 164.11 * FI_R - 422 * Level_L + 155.89 * Spread_L + \\
 & 408.66 * Level_R - 274.77 * Spread_R - 311.35 * \sigma_{LS} + 228.91 * \sigma_{RS} + 417.94 * LagValue + \\
 & 6.836 * X_L(1)
 \end{aligned}
 \tag{IV-15}$$

The analysis of the developed equation is found in Figure IV-24. The figure shows that the model has minimal outliers and an accuracy of 8 meters. The p-value of the model is less than 0.001 and therefore the model is statistically sound, and the null hypothesis is void. The r-

squared of the model is 98.23%, proving the high quality of the correlation. The correlation factor between the predicted, tested results and the real values is 0.99, proving a high correlation between the model and the phenomenon described by the model. The f-value of the model from Table IV-3 is calculated to be 934.89, which is a significantly greater value than zero and thus proves that the model is statistically robust and significant. Additionally, the utilized variables provide no impactful p-value, thus negating the null hypothesis.

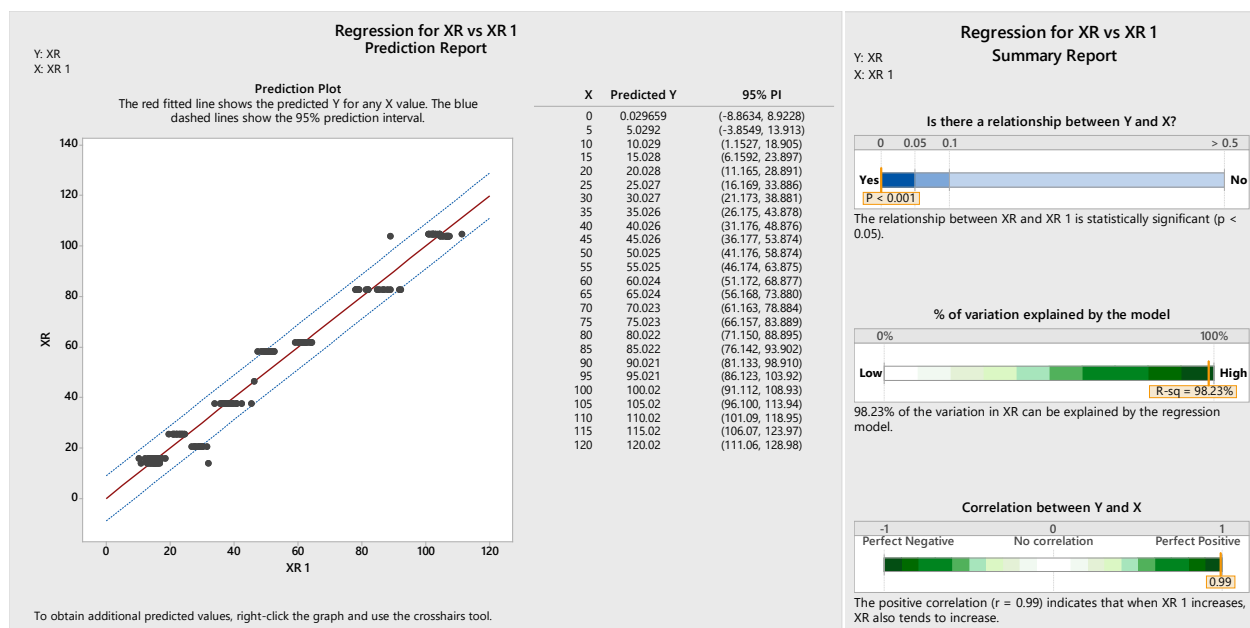


Figure IV-24: Acoustic $X_R(2)$ Regression Prediction Report

IV.3.2.3 Combined Pinpointing Models

Another concept for regression equations is to combine the models into a third new model using the equations in sections IV.3.2.1 and IV.3.2.2. The results of the two models are used for a new regression analysis against the real X_L and X_R values. The result of the regression are two new sets of equations that aim to form a relationship between the equations of each class to reduce the

noise and make the prediction more accurate. The models are dubbed X_{LF} and X_{RF} and are illustrated by equations IV-16 and IV-17. Equation IV-16 aims to set up a relationship between the output of equations $X_L(1)$ and $X_L(2)$. The equation also tries to develop a balanced state between the two equations such that it minimizes the error developed by both equations and gives a more stable final output.

$$X_{LF} = 0.676 + 1.0054 * X_L(1) - 0.0056 * X_L(2) \quad \textbf{(IV-16)}$$

In Figure IV-25, the aspects of equation IV-16 are studied. The model is displayed as a red line surrounded by the testing data used for model validation. The model has displayed a deviation accuracy of ± 3.8 meters to be 95% accurate in detection. Yet, the data points displayed in the figure show a much narrower window of prediction and a higher level of accuracy than the one reported by the software. Furthermore, the model has a p-value that is less than 0.001, which negates the null hypothesis and testifies for the health of the model. Finally, the model has a 99.49% value for r-squared, thus proving the significant level of correlation between the model and the variable of interest. To advance the model, the f-value of the equation is calculated and displayed in Table IV-3. The f-value is equal to 18000, which attests to the high level of significance of the model. Furthermore, $X_L(1)$ presents a high p-value in the model that is rectified by the addition of $X_L(2)$ to the model to make the p-value of the model less than 0.000.

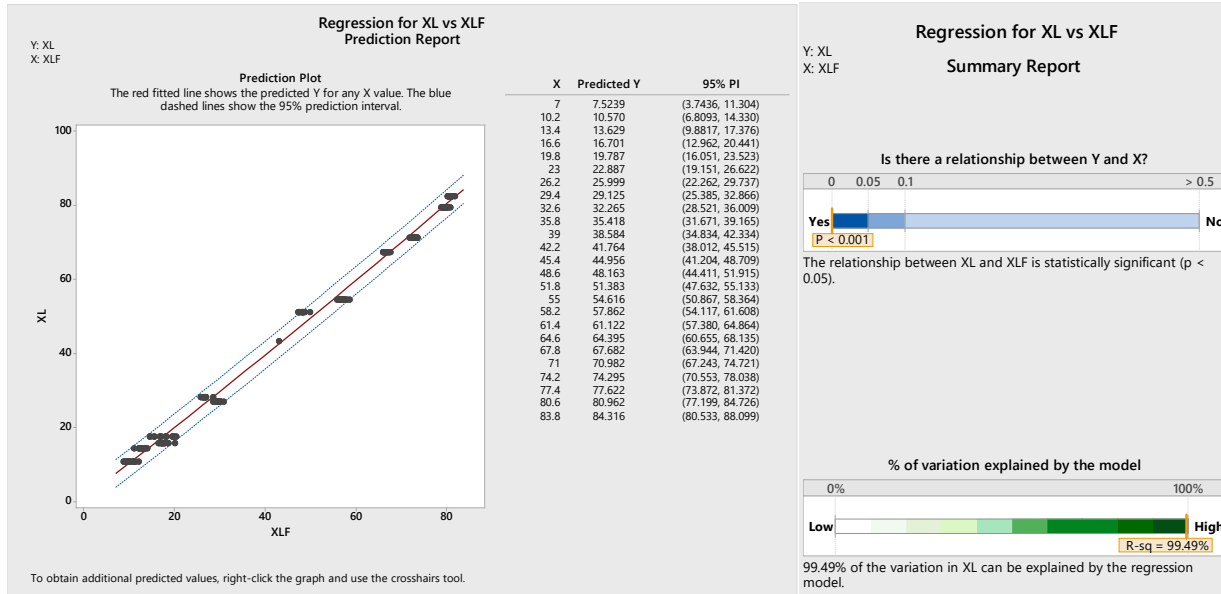


Figure IV-25: Acoustic X_{LF} Regression Prediction Report

In parallel, a combinatorial model for X_R is developed to mitigate the errors between the two previously developed models $X_{R(1)}$ and $X_{R(2)}$. The regression helps in developing a more comprehensible and straightforward equation that is much easier to realize and implement. The X_{RF} developed model is displayed in equation IV-17.

$$X_{RF} = 1.035 - 0.1792 * X_{R(1)} + 1.1689 * X_{R(2)} \quad (\text{IV-17})$$

The analysis of the equation can be studied in Figure IV-26. The equation, marked in red, is surrounded by the testing data points with a detection accuracy envelope of ± 8 meters. On the other hand, the values displayed by the testing show a much higher and cleaner accuracy in detection with much less deviation. The p-value produced by the model is less than 0.001 to negate the null hypothesis. Furthermore, the r-squared of the model is evaluated at 98.38%, thus displaying a highly correlated relationship between the variables and the studied outcome. The

conclusion is confirmed by having a correlation coefficient of 0.99 for the model. Table IV-3 displays the f-value of this model to be 5674.2, which is a highly significant value that confirms the results of r-squared and p-value. Additionally, throughout the development of the model, the constant value provides a high p-value for the model but is rectified by the addition of $X_R(1)$ and $X_R(2)$.

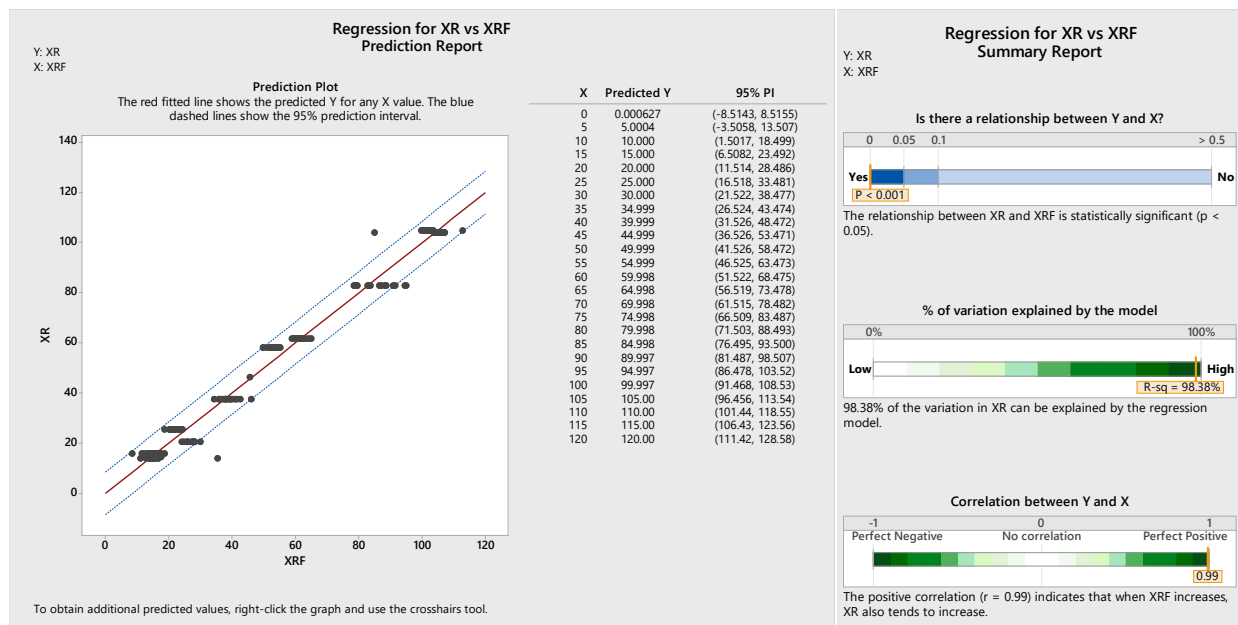


Figure IV-26: Acoustic X_{RF} Regression Prediction Report

IV.3.2.4 PX_L and PX_R Models

The final proposed conceptual model for the pinpointing of leaks within water mains monitored by acoustic loggers and listeners is the percentage model. The percentage model aims at developing an equation using the developed X_{LF} and X_{RF} and the distance between the two sensors, D , to determine a percentage value from the total distance. This model aims to mitigate the deviation to study if it can provide more accurate results in identifying the exact leak

location. The first model, illustrated by equation IV-18, represents the percentage of X_L from the total distance between the two sensors, D .

$$PX_L = 44.037 + 1.1226 * X_{LF} + 0.1312 * X_{RF} - 0.5536 * D \quad (\text{IV-18})$$

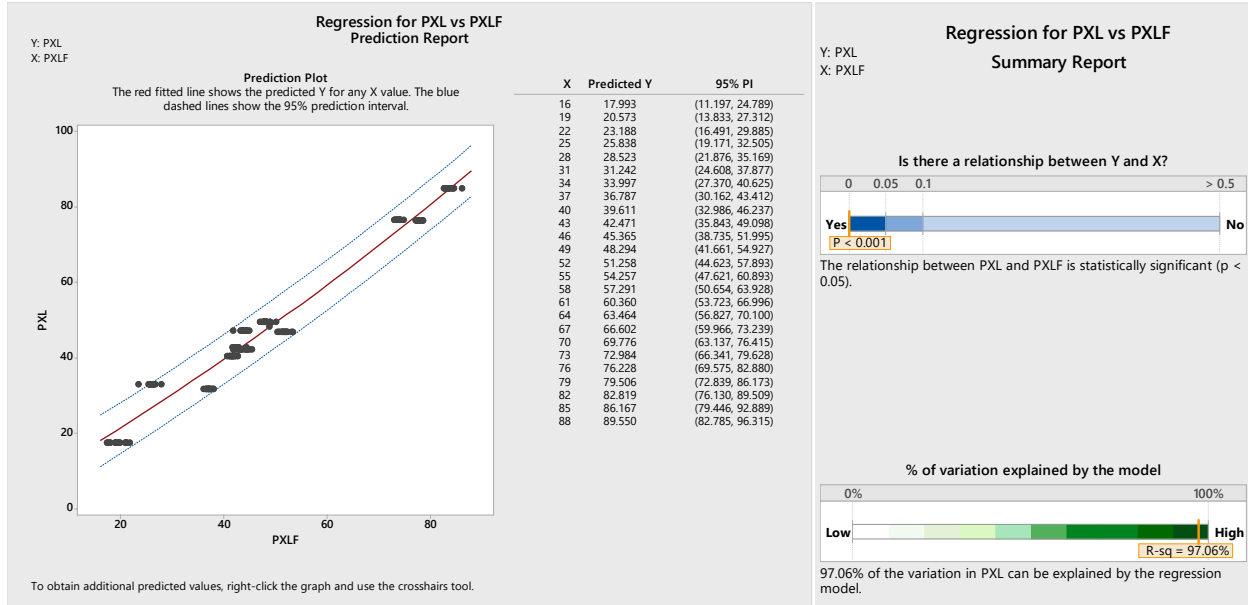


Figure IV-27: Acoustic PX_L Regression Prediction Report

The analysis of the model is summarized in Figure IV-27. The model is marked by the red line surrounded by black points that represent the testing data points. The average deviation of detection is displayed to be $\pm 6\%$ of the value. The big distance can create confusions as it can be highly accurate for short distance problem where the two loggers are near each other. On greater distances, a percentage model can suffer from great deviations. The p-value of the model is less than 0.001, thus negating the null hypothesis. Additionally, the model has a high correlation percentage represented by having a 97.06% value for r-squared. Finally, the f-value and progressive p-values of the variables can be viewed in Table IV-3. The f-value of this model is

estimated at 1980.91, which proves the statistical significance of the model. Furthermore, X_{RF} provides an impact on the p-value but the overall value is rectified with the addition of the total distance D . Complementarily, a model for PX_R is developed to assist the PX_L model. The variables are similar to the PX_L model as they include X_{LF} , X_{RF} and the distance between the two sensors, D . The output of the regression analysis is represented by equation IV-19 complementing the previously developed equation.

$$PX_R = 55.963 - 1.1226 * X_{LF} - 0.1312 * X_{RF} + 0.5536 * D \quad (IV-19)$$

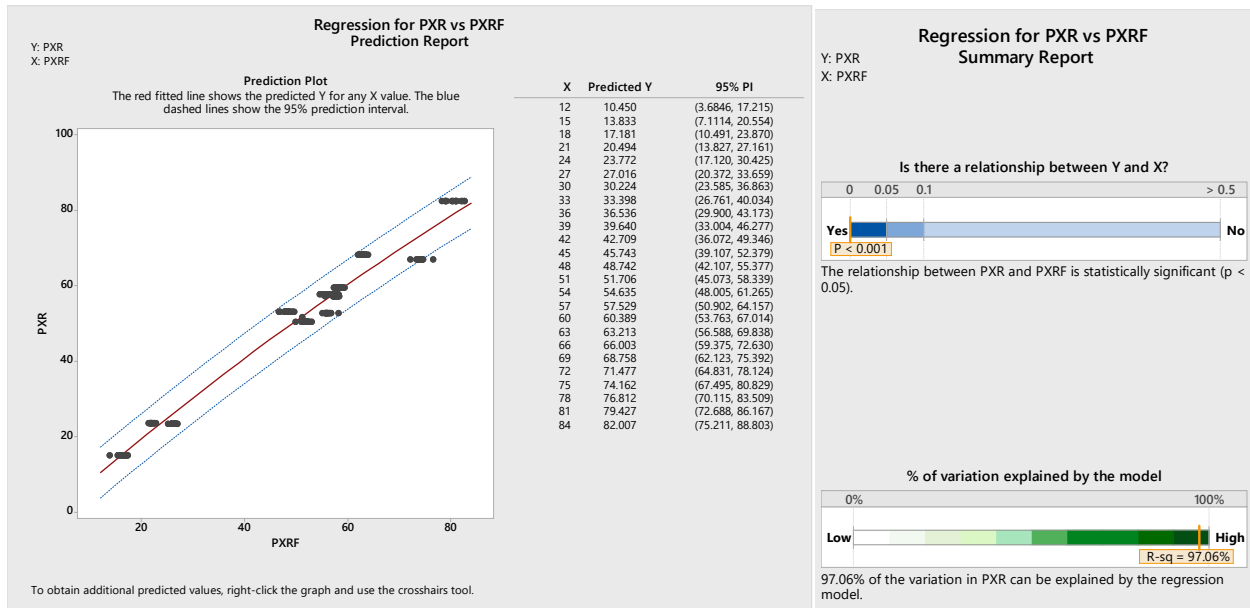


Figure IV-28: Acoustic PX_R Regression Prediction Report

The analysis of the model is highlighted by Figure IV-28 showing a deviation accuracy of $\pm 7\%$. The p-value of the equation is less than 0.001 to eliminate the null hypothesis and assure the soundness of the model. Finally, the r-squared of the model is high as it is estimated at 97.06%. Additionally, the f-value of the model in Table IV-3 is equal to that of the PX_L model at 1980.91,

confirming that the two models complement each other. Furthermore, X_{RF} provides an impact on the p-value but the overall value is rectified with the addition of the total distance D.

IV.3.2.5 Statistical Analysis of Developed Models

The developed statistical models require statistical analysis to validate their strength and applicability to the performed experiment and real-world applications. Table IV-3 has been extensively utilized in the previous sections to highlight the values and results of the statistical analysis of the developed models. Table IV-3 is comprised of six columns. The first column represents the code number of the equation under study. The second column represents the number of the equation in the text. The third column represents the respective r-squared percentage for the equation. The fourth column represents the overall p-value of the model. If the p-value is less than 0.001 and the r-squared value is closer to 100%, it is plausible to check for the fF-value. If the fF-value is greater than zero, the model is statistically significant. The higher the fF-value, the better it attests to the statistical significance of the model. The final column in the table displays the change in the p-value at each step of the model development and the change implied by each variable in the final version of the equation. Hence, the order of the variables listed within the table is of high significance.

Table IV-3: Advanced Statistical Analysis of Acoustic Leak Pinpointing Models

Code	Model Equation Number	R ² (%)	Model P-value	F-value	Coefficient P-value
1	IV-12	99.49	< 0.001	3029.13	<ul style="list-style-type: none"> • Constant = 0.000 • D = 0.000 • FI_L = 0.000 • FI_R = 0.000 • Level_L = 0.000 • Spread_L = 0.000 • Level_R = 0.000 • Spread_R = 0.000 • σ_{SL} = 0.000
2	IV-13	91.69	< 0.001	342.47	<ul style="list-style-type: none"> • Constant = 0.000 • Spread_R = 0.000 • Level_R = 0.000 • D = 0.000 • σ_{RS} = 0.000 • Level_R* σ_{RS} = 0.000 • Spread_R* σ_{RS} = 0.000
3	IV-14	97.08	< 0.001	2788.32	<ul style="list-style-type: none"> • Constant = 0.000 • Spread_R = 0.000 • Level_R = 0.000 • D = 0.000 • σ_{RS} = 0.037 • FI_L = 0.000 • FI_R = 0.000 • Level_L = 0.000 • Spread_L = 0.000 • σ_{LS} = 0.000 • Lag Value = 0.000
4	IV-15	98.23	< 0.001	934.89	<ul style="list-style-type: none"> • Constant = 0.000 • Spread_R = 0.000 • Level_R = 0.000 • σ_{RS} = 0.000 • FI_L = 0.000 • FI_R = 0.000 • Level_L = 0.000 • Spread_L = 0.000 • σ_{LS} = 0.000 • Lag Value = 0.000 • X_L(1) = 0.000

5	IV-16	99.49	< 0.001	18000	<ul style="list-style-type: none"> • Constant = 0.012 • $X_L(1) = 0.872$ • $X_L(2) = 0.000$
6	IV-17	98.83	< 0.001	5674.2	<ul style="list-style-type: none"> • Constant = 0.078 • $X_R(1) = 0.000$ • $X_R(2) = 0.000$
7	IV-18	97.06	< 0.001	1980.91	<ul style="list-style-type: none"> • Constant = 0.000 • $XLF = 0.000$ • $XRF = 0.041$ • $D = 0.000$
8	IV-19	97.06	< 0.001	1980.91	<ul style="list-style-type: none"> • Constant = 0.000 • $XLF = 0.000$ • $XRF = 0.041$ • $D = 0.000$

IV.3.2.6 Comparative Validation of Developed Models

Following the development of the four types of models are illustrated in the previous sections. Those models should be validated against real data extracted from the field. Therefore, a leak has been discovered by the city of Montreal on Viger street and its data is collected in the form of sound signals. The sound signals are analyzed as explained earlier and their data is input into the developed models to analyze their results. Furthermore, the water service agency in the city of Montreal has used geophones and hydrophones to verify the exact location of the leak from the sensors that detected its existence. Table IV-4 presents the results of the findings from testing the models against the real leak. The X_L location of the leak is 84.9 meters and the X_R distance of the leak is 154 meters. Accordingly, the results of each class of models are recorded to test their extent in detection. The initial models have the highest levels of metric deviations with 3.2 meters of deviation in the left direction and 16.7 m deviation in the right direction. On the other hand, the second set of models has performed better than the initial models with a 20-centimeter deviation in the left direction and no deviations in the right direction. The combined models do well in mitigating the deviations of the initial model by complementing it with the results of the

second model. The deviations are at 1.9 meters from the left direction and 3 meters from the right direction. Finally, the percentage models have the most significant deviation due to the size of the distance between the two sensors with 11% deviation in each direction. It is safe to conclude that the second set of models are capable of accurately detecting leaks, yet it is highly recommended to use all the available models to have a sense of the leak situation that is being studied.

Table IV-4: Viger Street Leak Validation of Leak Pinpointing Models

	Actual Location	Initial Models	Second Models	Combined Models	Percentage Models
X_L(m)	84.9	81.7	84.7	83	24%
X_R(m)	154	170.7	154	151	76%
Deviation X_L(m)	----	-3.2	-0.2	-1.9	-11%
Deviation X_R(m)		+16.7	0.0	-3.0	+11%

CHAPTER V: REPAIR PRIORITIZATION MODELS

V.1 Repair Prioritization Model Development Methodologies

The main aim of this section is to develop novel approaches for analyzing multiple deteriorating events that coexist at the same time. After analysis, the events are then prioritized based on a set of selection criteria that is defined by the user to maximize the expected benefit against the spending on the repair events. This chapter fulfills the third objective of this thesis and that is to build an optimized model for leak repair prioritization.

V.1.1 The Lazy Serpent Algorithm

Municipalities deal with more than one leak at the same instance and hence face the question of “which leak has the priority over the rest to repair?” This section proposes a new prioritization algorithm with a three-dimensional approach to tackle the leak prioritization problem. Figure V-1 presents a preliminary model for the Lazy Serpent Algorithm that is currently under development. To develop the Lazy Serpent Algorithm, research is needed in terms of simulation techniques and optimization algorithms. The primary purpose of having a background simulation technique is to predict the behavior of the leak (event) as time progresses, whereas optimization algorithms are utilized to determine the best path for repairing the maximum number of leaks with the minimum amount of damages or effects as predefined by the user. The impact estimator is a smart decision-making merger between simulation and optimization that allows the serpent (resource or repair team) to select the least problematic event; this approach is currently still under development. The Lazy Serpent Algorithm can tackle the time consumption problem in other prioritization algorithms by immediately going towards a user-oriented near-optimal solution. This is done by the algorithm’s capability to immediately start constructing a solution that optimizes the priorities provided by the user. Furthermore, the three-dimensional definitions

of the Lazy Serpent Algorithm allow municipalities and other organizations that require prioritization of independent events to define and select the criteria that most matter and define their changes. Thus, the algorithm creates a field for analysis that targets only their needs based on their available data. Additionally, the Lazy Serpent Algorithm is created to solely solve the prioritization problem. Therefore, it is much simpler to use and implement than other algorithms requiring a lot of definitions and setups. Figure V-2 summarizes the components of the Lazy Serpent Algorithm. The algorithm contains two main compartments, boundaries and inputs. The input compartment is composed of two partitions, that are inputs on the level of events and inputs on the level of resources (serpents).

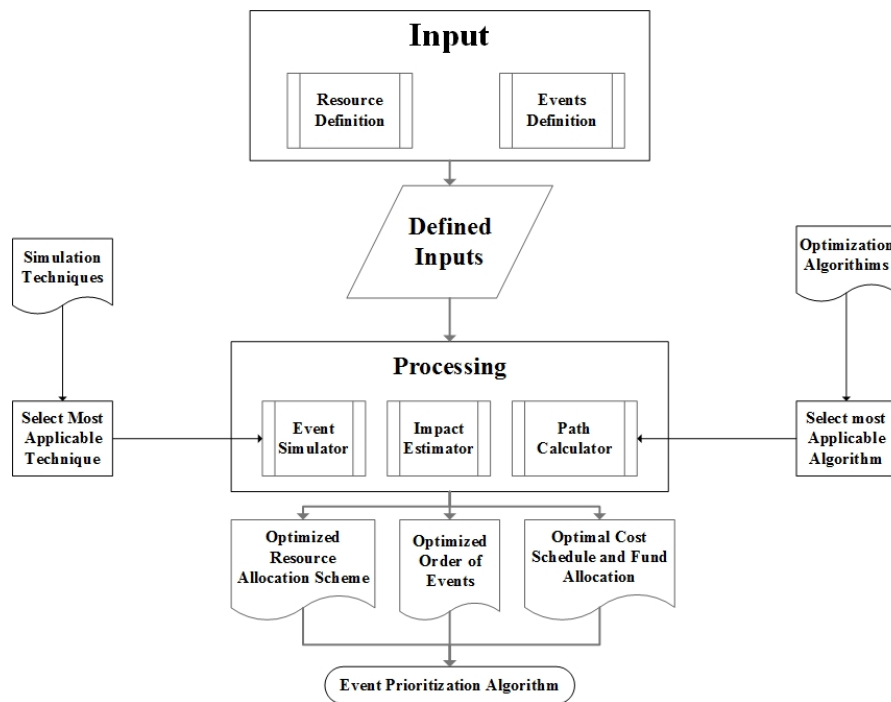


Figure V-1: Lazy Serpent Algorithm Preliminary Model

Each partition has its type of inputs. There are four main types of event inputs and they are (1) Basic Events (BE), (2) Serpent-Specific Events (SSE), (3) Multi-Serpent Events (MSE) and (4)

Any-Serpent Events (ASE). MSEs are divided into two types, (1) Simultaneous (SMSE) and (2) Consecutive-Layers (CLMSE). Accordingly, the input of resources (serpents) is characterized by three primary attributes as (1) Behavior, (2) Color and (3) Consumption Rate. Behavior is divided into two behaviors as (1) Establishing Behaviors and (2) Returning Behaviors. Establishing behaviors can be either independent or pack mode behaviors.

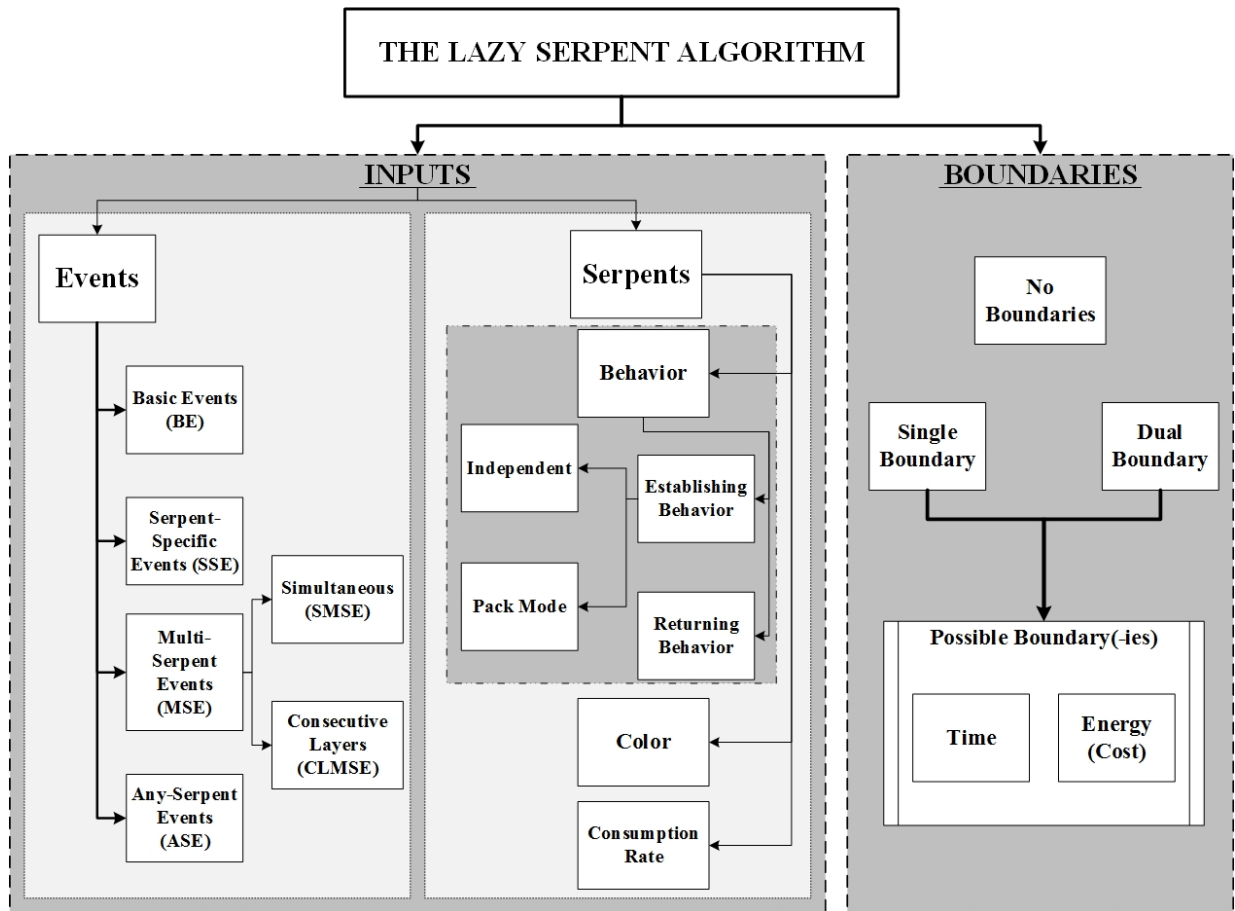


Figure V-2: Lazy Serpent Algorithm Components Summary

The second compartment is boundaries, which are the constraints of the algorithm. The Lazy Serpent Algorithm has two main constraints as (1) Time and (2) Energy (funds or money). Each of those two constraints can be used separately or they can be used together. Another option is to

avoid the use of constraints to identify the ideal solution. The aspects above will be defined and explained further throughout this section. The current version of the algorithm is crude and requires further development on multiple aspects. Yet, it can be summarized as follows:

V.1.1.1 Definition

The Lazy Serpent Algorithm simulates the motion of hungry serpents (snakes) throughout a field trying to eat hopping rabbits or rolling eggs (moving events) in the 3D field. The serpent is lazy and thus it tries to follow the most profitable path with the least effort. Besides, when a serpent deals with an event, the serpent will not be available within the specified amount of the time required to fix the leak (event). The algorithm relies on representing resources as serpents and on upcoming events as moving eggs or hopping rabbits. The events are expected to have an initial condition and equations that represent their motion through the three dimensions of the field. The three dimensions are specified by the user with a minimum and a maximum and they are preferred to be equally weighted and uniform in direction (decreasing or increasing from maximum to minimum). Each event is expected to move with each of the three dimensions X, Y, and Z with a specified dimension equation $f(x)$, $f(y)$ and $f(z)$, as predetermined by the user. Each event is expected to have a weight, which represents the amount of time required to complete this task/event. During this time, the resource (serpent) attached to this event is considered busy and unable to function until the weight time passes.

V.1.1.2 Inputs

To operate the Lazy Serpent Algorithm, a set of inputs is required to set up the three-dimensional space. Primarily, the events are distributed based on the three axes of importance specified by the

user, with their specific range preferably 0 to 10. The inputs are classified on three levels events, resources and boundaries. The events in this specific case are discovered leaks and their respective information, as they are required by the algorithm and are further explained in section V.1.1.2 part A. Resource input, also referred to as serpent input, indicates the number of available resources with their respective specific algorithm information. In the particular example given, it refers to the number of available repair teams as highlighted in section V.1.1.2 part B. The third type of input is boundaries; however, boundaries are not a necessity for the operation of the lazy serpent and thus the algorithm can run without any constraints to find the optimal solution regardless of the time or the resources required. However, constraints can be added to maximize or minimize the particular aspect the user wishes to tackle. For instance, a time constraint can be added to let the algorithm maximize the number of repairs in the specified limit. Boundaries are further explained in section V.1.1.2 part C.

A. Event Input

An event is simply an occurrence that has a certain state at the time it takes place. Thus, an event within the Lazy Serpent Algorithm is assumed to have an initial condition on all three axes of assessment (x_0 , y_0 , and z_0), in addition to the change of condition functions in each axis ($f_{(x)}$, $f_{(y)}$, and $f_{(z)}$). The change equations are assumed to be a function of the same unit of time. In addition to the initial condition and the expected change in condition through time, an event will have two other parameters that identify the time expense and the financial expense and they are t_n and c_n . T_n indicates the amount of time needed for the complete disposal of the event and c_n indicates the total cost needed for handle the event completely. Therefore, an event A can be represented as A (x_0 , y_0 , z_0 , $f_{(x)}$, $f_{(y)}$, $f_{(z)}$, t_n , c_n). The definition is further described in Equation (V-1), where the

initial condition is part of the motion equations $f(x)$, $f(y)$, and $f(z)$. The equations are composed of the initial condition x_0 , y_0 and z_0 along with their relative variation through time $g(t)$.

$$A = \begin{cases} f(x) = x_0 + g_x(t) \\ f(y) = y_0 + g_y(t) \\ f(z) = z_0 + g_z(t) \end{cases}, t_n, c_n \quad (\text{V-1})$$

For example, let us assume an event A with initial conditions (6,8,4) and a change in time g (-0.1t, -0.4t, 0) and a t_n of 8 days and a c_n of 100,000\$. Therefore, $f(x)$, $f(y)$ and $f(z)$ would be (6-0.1t, 8-0.4t, 4). In case we want to determine the condition after four days of the initial investigation, we replace t with 4 and the results would be (6-0.1(4) = 5.6, 8-0.4(4) = 5.4, 4) and A_4 would be (5.6, 5.4, 4). Throughout the development of this algorithm, multiple conditions have been identified that require the existence of multiple types of events to deal with each type of real-life event appropriately. The identified event types are enlisted as follows:

i - Basic Events (B):

The first type of events is a basic event. A basic event is an event that has no specific nature or a specific approach such as delivering a simple item or refilling a gas tank. Regarding leaks, it can be iterated as a simple leak that any available team can easily dealt. Thus, a basic event would have no specific color and can be handled by any resource as illustrated in Figure V-3. The drawing illustrated in Figure V-3 shows that the basic Event A is capable of getting eaten by any of the three serpents on the basis of first come first served. A basic event is represented as a colorless circle with a dotted circumference. The value eight located within event A represents the weight t_n that presents the time required to complete event A. Equation (V-2) presents the mathematical and logical presentation of basic events. For a devouring to occur (e.g. event handling or leak repair for example), the event must be at the shortest distance from any serpent

and that serpent must be idle. If the two conditions are met, the event begins to be devoured and thus, in real life, the leak repair begins.

$$(d = \min \wedge \text{serpent} = \text{idle}) \Rightarrow \text{devour} = \text{true} \quad (\text{V-2})$$

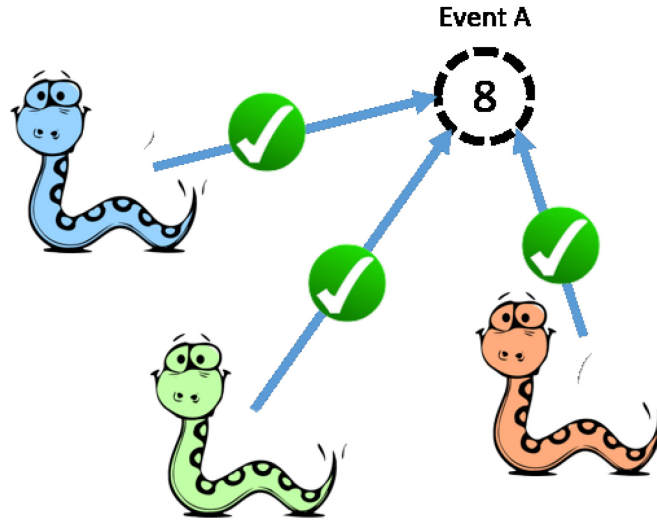


Figure V-3: The Reaction of Serpents to Basic Events

ii - Serpent Specific Events (SS):

The second type of events are those that can be solved by only one type of resource. In this algorithm, they are dubbed as “Serpent Specific Events” and are represented by a specific color. Thus, only the matching serpent can handle this event and any other type of serpent would merely ignore it. The illustration in Figure V-4 displays an event B colored in blue. Thus, the orange and green serpents are unable to react event B and only the blue serpent can react event B and complete it, whereas the rest stay idle even if they are available. A specific event is represented as a full circle filled with the color of the necessary type of serpents along with the time weight t_0 . Mathematically, serpent specific events can be illustrated using Equation (V-3). The equation shows that, to have an event interaction between serpent and event, three

conditions must be met. The distance conditions and the idle condition are the same from the previous section. The new condition for this type of events is color. The color of the event ($color_e$) and the color of the serpent ($color_s$) must match precisely to start the interaction.

$$(d = \min \wedge \text{serpent} = \text{idle} \wedge \text{color}_e = \text{color}_s) \Rightarrow \text{devour} = \text{true} \quad (\text{V-3})$$

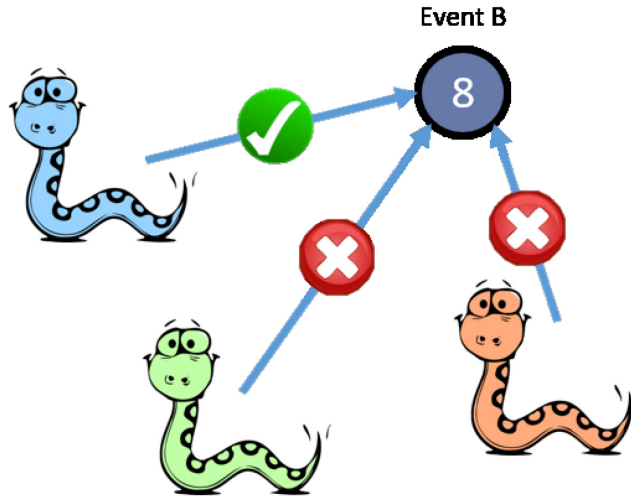


Figure V-4: The Reaction of Serpents to Serpent Specific Events

iii - Multi-Serpent Events (MS):

In real life, not all events or actions require one resource or team to be completed. As a result, multi-serpent events have been developed. Multi-serpent events are events that require more than one resource (serpent) to complete and they are divided into two types:

a – Simultaneous: a simultaneous event is an event that requires the existence of two specific resources or more at the same time to be completed. Once commenced, all resources will suffer the same delay (t_n) required for the event but the cost of the event (c_n) will be one regardless of the number of used resources. Further costs as the operational costs of each serpent are referred to as energy consumption; this concept will be later explored in Section III.2 part B.

Figure V-5 displays how serpents react in the presence of a simultaneous event. Event C is not tackled when either the blue or green serpent is available; then, event C is tackled and completed. A simultaneous event is drawn as a circle that is equally divided by the colors representing the required serpents along with the amount of the required time t_0 .

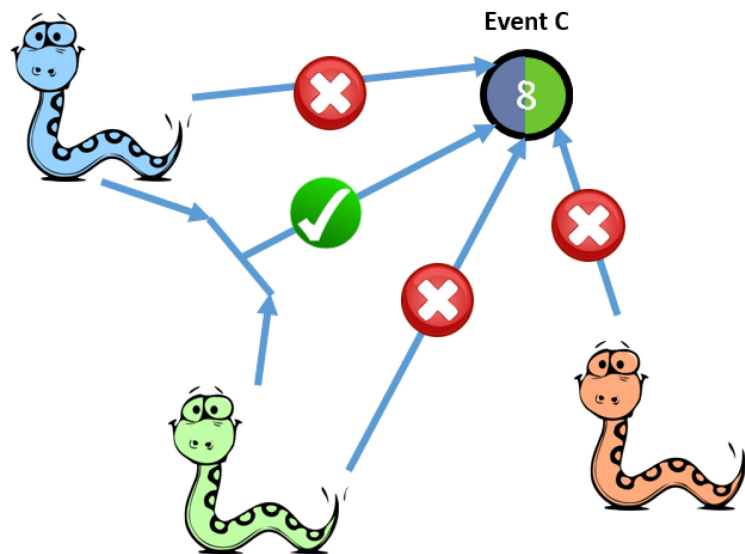


Figure V-5: The Reaction of Serpents to Simultaneous Multi Serpent Events

On the level of mathematical representation, Simultaneous Multi Serpent Events (SMSE) can be represented as a decision tree with three main decision criteria that must coexist at the same time. Each criterion has two primary objectives to be validated for it to be true. When all three criteria are valid, then the execution can be carried out. Equation (V-4) illustrates this idea by highlighting the three required conditions to carry out the assessment leading to decision making. The first criterion is distance, where the two serpents are checked for their proximity from the event. The second is idleness and the last criterion is color matching.

$$(d_1 \ \&\& \ d_2 = \mathbf{min} \wedge s_1 \ \&\& \ s_2 = \mathbf{idle} \wedge cs_1 \ \&\& \ cs_2 = \mathbf{ce}_1 \ \&\& \ ce_2) \Rightarrow \mathbf{devour} = \mathbf{true} \quad (\mathbf{V-4})$$

b – Consecutive-Layers: Although prioritization algorithms do not normally deal with events with interdependencies, a certain type of events was noted with certain interdependencies that require being tweaked into the Lazy Serpent Algorithm. Consecutive-Layers events are those that require more than one resource to be completed. The required resources must be available in a certain order. Therefore, consecutive-layers events are considered layered events that change their color after a serpent completes a layer to become another task that requires another resource. Figure V-6 illustrates this concept by displaying event D that has a green outer layer and a blue sublayer. Event D will be treated as a green event until a green serpent (resource) is available to complete it. When completed, event D becomes a blue event and waits for an available blue serpent to be finalized. The blue serpent cannot tackle event D before the green serpent. Only after the green serpent is done, can the blue serpent tackle event D. In terms of cost, a consecutive-layers event will not be charged until completion, i.e. until all layers are resolved. In this figure, the first layer requires four days to be performed after which the event will become a blue event and require eight days to be finalized. On the level of coding and mathematical representation, consecutive-layers events are considered serpent specific events based on their initial layer as in Equation (V-3). However, after the primary interaction is completed, the devour command will trigger another command as in Equation (V-5), which is the transform command. The transform command transforms the event into another event after the serpent-event interaction. The equation displays the transformation of an event A_i from its first layer (A_{i1}) to its second layer (A_{i2}).

$$\text{devour} = \text{true} \Rightarrow \text{transform} (A_{i1} \text{ to } A_{i2}) \quad (\text{V-5})$$

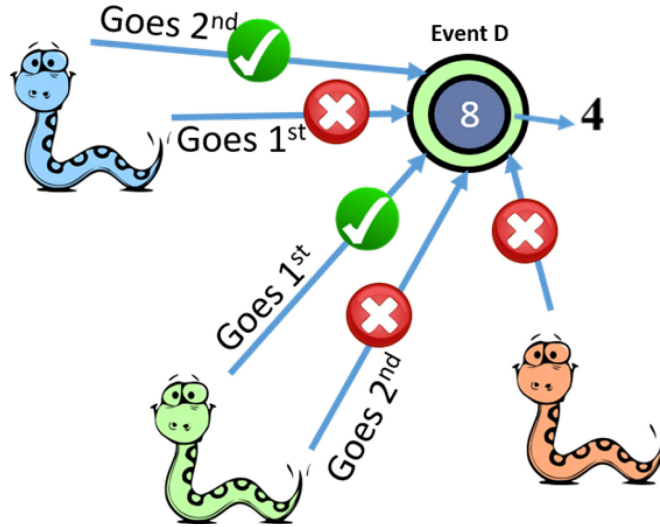


Figure V-6: The Reaction of Serpents to a Consecutive Layers Multi Serpent Events

iv - Any-Serpent Events (AS):

Another type of events in the field is characterized as those that can be performed by a specific pool of resources. Thus, the difference between a basic event and an any-serpent event is that any existing resource can handle basic events, whereas any serpent events are handled by any serpent from a predefined group of resources. Thus, as in Figure V-7, an any-serpent event is displayed as an equilateral triangle that is colored equally by the serpents (resources) capable of handling it. The figure shows event E is colored half orange and half blue, which means only the blue and the orange serpents can deal with event E. As a result, whichever of the two serpents is available first handles this task and has to spend the amount of time t_n , which is 8 days in the case of Figure V-7, that is required for completion. At that time, the serpent is categorized as digesting (unavailable resource), which means busy and unable to move or leave the location of event E.

Regarding mathematical representation, any-serpent events are similar to serpent specific events except for the matching condition where the closest serpent must belong to a pool of colors. As in Equation (V-6), the serpent is first checked for proximity, then availability and idleness, and finally if it belongs to a predefined set of colors.

$$(d = \min \wedge \text{serpent} = \text{idle} \wedge \text{color}_s \in \{c_{e1}, c_{e2}\}) \Rightarrow \text{devour} = \text{true} \quad (\text{V-6})$$

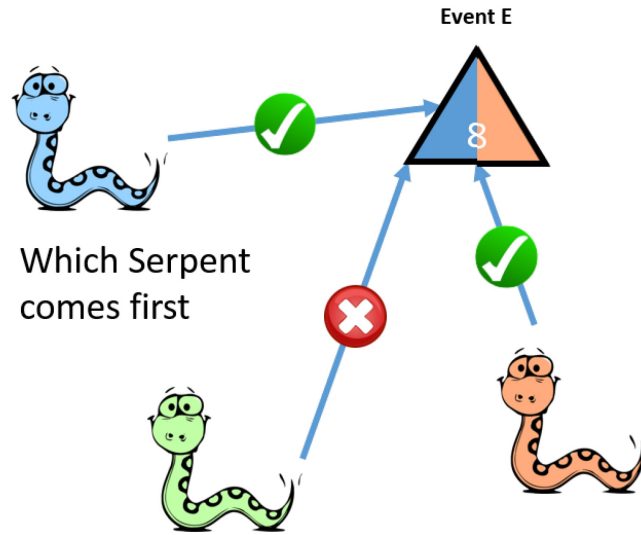


Figure V-7: The Reaction of Serpents to Any Serpent Events

B. Resource Input (Serpent Definition)

All serpents are governed by one single motion or action type called behavior and each serpent is then defined based on color and consumption as the two main variables. Regarding color, each serpent can have only one color, which represents the type of action the resource (serpent) could perform. However, consumption signifies the amount of energy resource consumed per time (t_n) by the resource. During the time required by the resource t_n , the serpent is unable to move and is referred to by the term “digesting” (occupied) and stays at the location of the event until the

process is complete. Beyond the completion of an event, the serpents will react per a predefined behavior that governs the behavior beyond the task completion.

The overall behavior defines the way the serpent behaves after devouring an egg (completing an event) and these reactions are divided into two main categories:

i – Returning behavior, assuming that all serpents will return to their headquarters, i.e. the point of origin specified by the user, which is default set at (0, 0, 0). Therefore, whenever a serpent (resource) finishes a task, the serpent will return to the origin point to select the next best event to tackle. Figure V-8 displays this behavior as all serpents first set out to their targets and, after completing each task, the serpents shall return to the point of origin. Since the green serpent had the event with the lowest time consumption, it returns to the origin first and it immediately sets on to go to a new target if it exists, whereas the orange serpent would be the last to return to the origin point due to having its event requiring the most time.

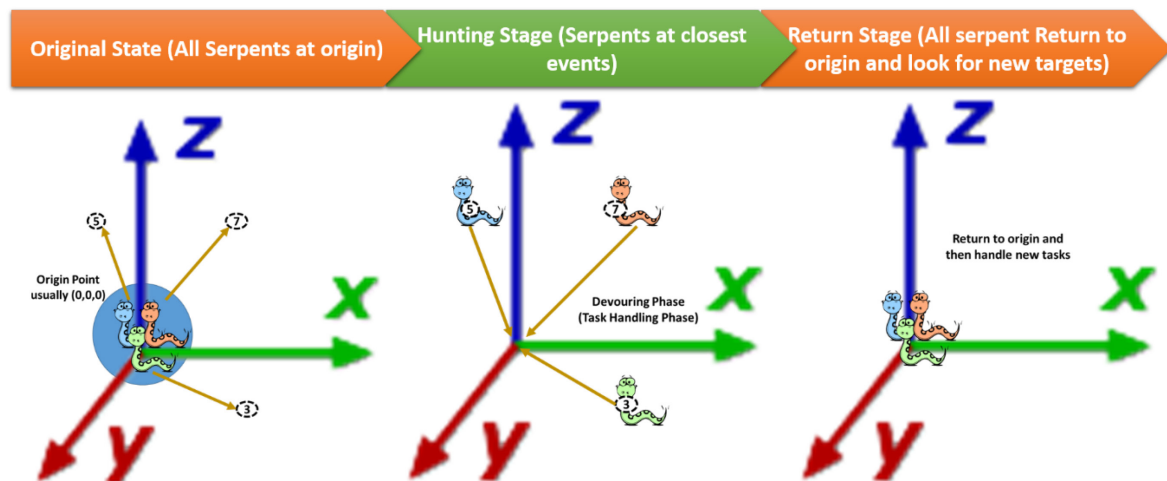


Figure V-8: Returning Serpents Schematic Explanation

ii – *Establishing behavior*: In practice, not all resources return to their starting point once they finish the task they are assigned. Rather, the resource would search for the problem nearest to its location to tackle it. Thus, the concept of establishing serpents or establishing behavior was developed to simulate the reality on the field. Establishing serpents can establish new origins as they are on their paths to determine the next best solution. Establishing serpents are divided into two main categories, independent and pack mode, and they are defined as follows:

a - *Independent*: In this category, it is assumed that each resource selects the location of the finished task and establishes its respective location as a new point of origin. Immediately after setting the new point of origin, the serpent moves to select the event location closest to the newly established origin and solve it. After solving the second event, the serpent establishes the new location as a new origin point and this process goes on. Figure V-9 illustrates the approach with more clarity. As highlighted, all serpents start at the origin point (0, 0, 0) and then move towards their targets. Later, after completing their tasks, each serpent has its origin point and it selects its next target accordingly.

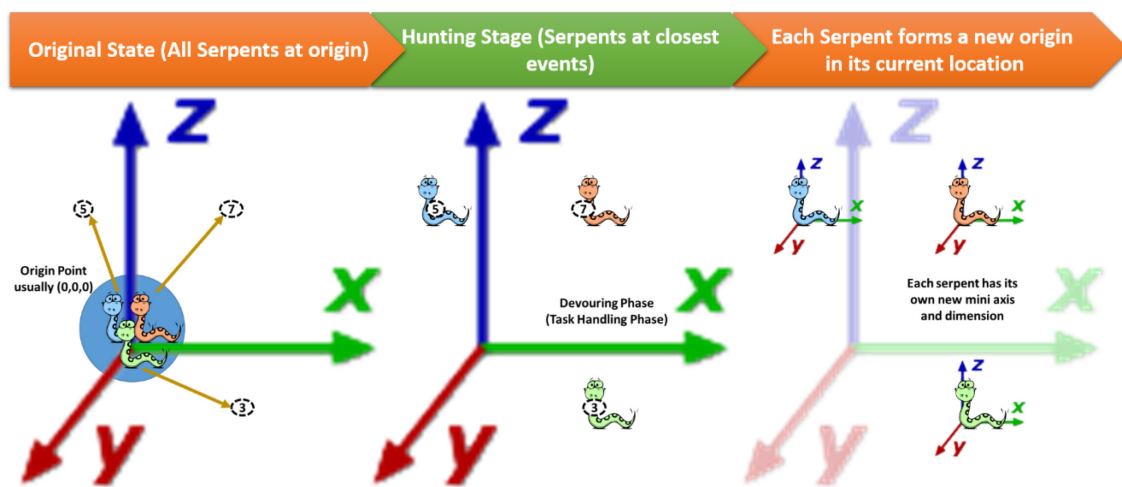


Figure V-9: Independent Serpent Behavior

b - *Pack Mode*: In some situations, a moving resource will be referenced as a strategic point or a moving command center that other units have to return to when they are done. Thus, a pack mode behavior requires the central resource to be selected and identified. The chief resource becomes the origin point as it moves and all other serpents have to return to the current location of the chief serpent (central resource) before going onto their next target. Figure V-10 displays this behavior by showing the chief blue serpent at the origin point before each serpent embarks on its task. After completion, all serpents return to the blue serpent's current location before heading off to their new tasks.

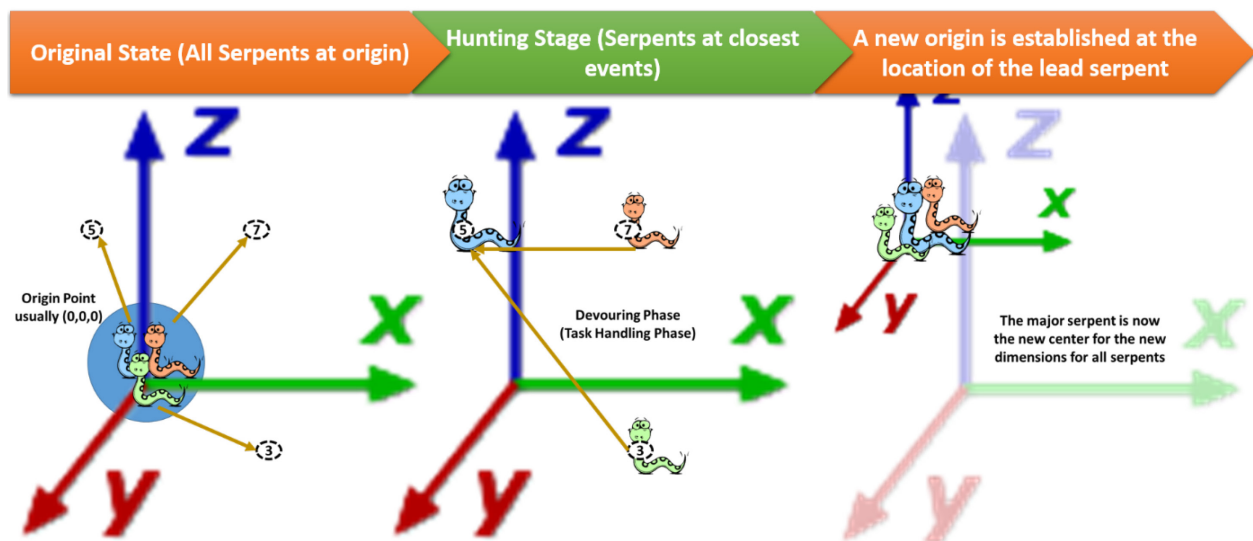


Figure V-10: Pack Mode Serpents Behavior

C. Boundaries

Boundaries are the third factor of setting the lazy serpent field. Although boundaries are not a must, they are critical to the decision-making process as they represent the constraints that the decision maker, e.g. municipalities and project managers, have at hand. The primary two

constraints for the Lazy Serpent Algorithm are time and energy (cost). Time as a constraint represents the total amount of time consumed by the overall algorithm and energy represents the cost consumed by the project. However, energy is a general term, too, as money might not be the only resource of concern that is consumed by the project.

Based on what is discussed above, the governing factor is the relationship between benefit and cost. This relationship is predetermined based on the type of the problem. For the problem of leak repair, all leaks are assumed to be deteriorating and therefore the faster they are repaired, the better it is for saving precious water resources and to minimize their impending damage. Based on this assumption, the benefit of an event is the collected value of its current point. For example, an event H is repaired when it reaches the point (4,7,6) in the 3D space. Therefore, its benefit is $b_{tH} = 4+7+6 = 17$. Further illustration is provided by Equation (V-7), where the benefit at termination b_{tH} is equal to the summation of the end of repair point's coordinates in the 3D space.

$$b_{ti} = x_{ti} + y_{ti} + z_{ti} \quad (V-7)$$

However, the cost is the summation between the cost needed (c_{ni}) by the event and the resource operational costs calculated by multiplying the time needed for repair (t_{ni}) by the cost rate of the resource (r_{cs}). The illustration of this concept can be found in Equation (V-8), where each event has its own overall cost.

$$C_{oi} = c_{ni} + (t_{ni} * r_{cs}) \quad (V-8)$$

Based on the two event equations, the total benefit of a solution is the summation of all the benefit values of all the events that comprise the selected solution as displayed in Equation (V-9). The equation shows the approach to determine the total benefit (B) of a solution by calculating the respective benefits conserved for each event that is selected for repair. Letter n refers to the total number of events within the solution.

$$B = \sum_{i=1}^n b_{ti} \quad (\text{V-9})$$

Similarly, the total cost of a solution is the summation of the costs of all the events comprising this solution as in Equation (V-10), where the total cost (C) is the sum of all the events in the solution whose number is n . Moreover, the benefit to cost ratio is assumed to be equal to the division of B from Equation (V-9) over C from Equation (V-10). This division represents the value that was prevented from deteriorating over the expenditure spent in the repairs.

$$C = \sum_{i=1}^n C_{oi} \quad (\text{V-10})$$

Thus, the following approaches are identified:

- i- Free/Unbound: The primary assumption in the unbound setup is that energy (money) and time are abundant and can be spent openly. The field of algorithm has no constraints and thus the algorithm runs until all events are completed and solved. Furthermore, the algorithm tries to figure out the best path with the lowest time and energy consumption. Equation (V-11) illustrates the optimization goal of this mode and that is to find the maximum benefit to cost ratio for all possible orders.

$$\max\{\frac{B}{C}\} \quad (\text{V-11})$$

- ii- Time Bound: In this approach, it is mainly assumed that energy (fund) is abundant but time is scarce. Thus, a time limit (T) exists, forcing the algorithm to stop after the running time (t_{run}) reaches the time limit (T). Thus, the main aim of algorithm is to maximize the benefit to cost ratio within the time limit regardless of cost. In Equation (V-12), the previously mentioned conditions and goals can be summarized by an optimization equation.

$$\max\{\frac{B}{C} : t_{run} \leq T\} \quad (\text{V-12})$$

- iii- Energy Bound: The exact reverse of time boundary is an energy boundary (E). The main assumption is that time is abundant and holds no constraints whereas funds are scarce and must be efficiently allocated. Therefore, the algorithm tends to maximize the number of tasks performed for a specific amount of energy (funds). Equation (V-13) displays the approach of maximizing the benefit to cost ratio for a maximum allotted energy or fund (E) that the solution expenditure or cost (e_{run}) must never exceed.

$$\max\{\frac{B}{C} : e_{run} \leq E\} \quad (\text{V-13})$$

- iv- Dual Boundaries: The main condition under this boundary is that both energy and time are scarce. Thus, the algorithm tends to maximize the number of tasks performed within the limits of time and energy as displayed in Equation (V-14).

$$\max\{\frac{B}{C} : t_{run} \leq T, e_{run} \leq E\} \quad (V-14)$$

The optimization equations above can be summarized in Equation (V-15). The main objective is to maximize the benefit of the incurred expenses, taking into consideration a specific timeline and overall budget. One or both constraints can be removed depending on the user's needs.

$$\begin{aligned} \max \quad & \frac{B}{C} \\ \text{s.t.:} \quad & \\ & t_{run} \leq T \\ & e_{run} \leq E \end{aligned} \quad (V-15)$$

Another field setup is the threshold. A threshold in the Lazy Serpent Algorithm is a set of values on each plane that represent a borderline the user, e.g. municipalities, prefer not to cross at all costs. A threshold is made of 3 planes with specific values along with the intersection with the origin planes. It is not necessary that all axes have thresholds, it is possible to have all or none. A threshold represents the worst condition from the eyes of the user and thus if an event crosses a threshold, the event has crossed into an undesirable condition. Once an event or multiple events cross the threshold, the Lazy Serpent Algorithm notifies the user that there might be a need to increase the number of resources to prevent further events from crossing the threshold.

D. Proposed Solution: Inverse Pyramid

To solve the lazy serpent problem, a greedy solution is proposed. The solution imagines the 3D space events in a planar space that consists of the points of distance calculated previously. The first event to be selected is the event with the lowest distance since it fits the budget and time

constraints, as in Figures V-11 and V-12. Figure V-11 shows that the algorithm starts with an initially assigned population of all the events under analysis. Whenever an event is contained, a new plane is created without the handled event. The selected mode of analysis is then performed on the new plane and the cycle is repeated until all constraints are met. Hence, the name “inverted pyramid” solution, where the solution starts with the complete population at first and starts creating new populations that are smaller than the earlier populations due to the eliminations process. The selected event is then removed from the plane and is replaced with infinity and a new plane is formed with an updated time that adds the repair time taken to repair the event. The solution would continue to run and eliminate solutions until it hits either the budget or the time constraint.

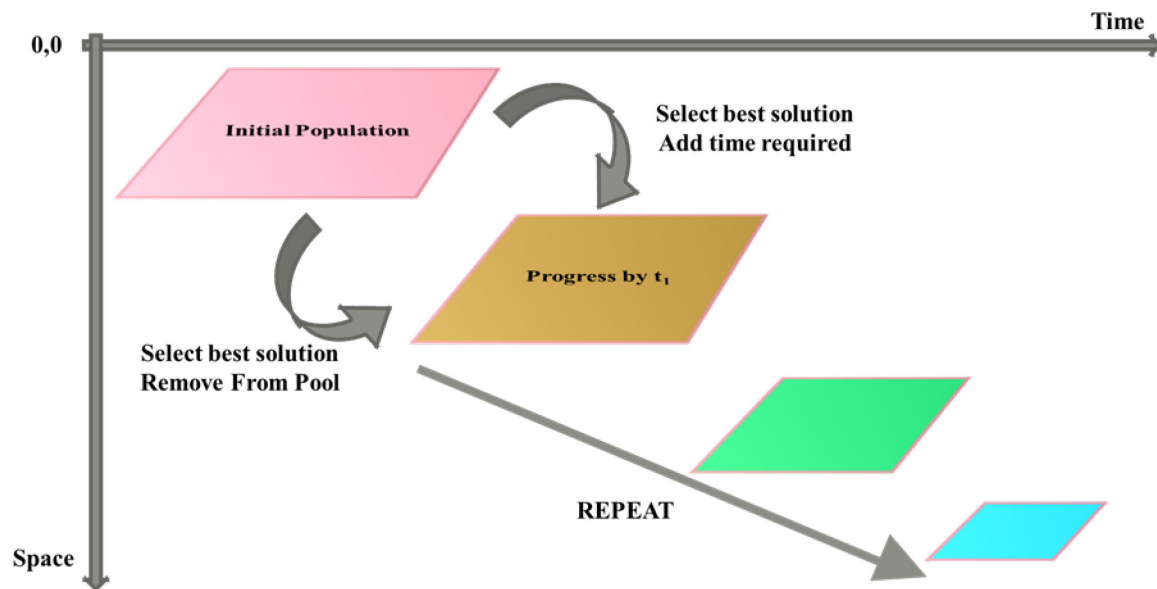


Figure V-11: Inverse Pyramid Solution Approach

Figure III-27 further illustrates the form of updating the equations of the solution. Initially, e_{Run} is equal to zero and, after initialization, the first event is selected. If the event meets the repair

criteria, then it is repaired and its cost is added to e_{Run} , which continues to update itself based on this criterion until it meets the budget constraint or finds no more available events for repair. Similarly, time is updated with the same fashion using the variable t_{Run} , which keeps on selecting and updating events simultaneously with e_{Run} . If one event does not meet the criteria of both equations, then the event is dropped and the code would try to find another event that fits the criteria.

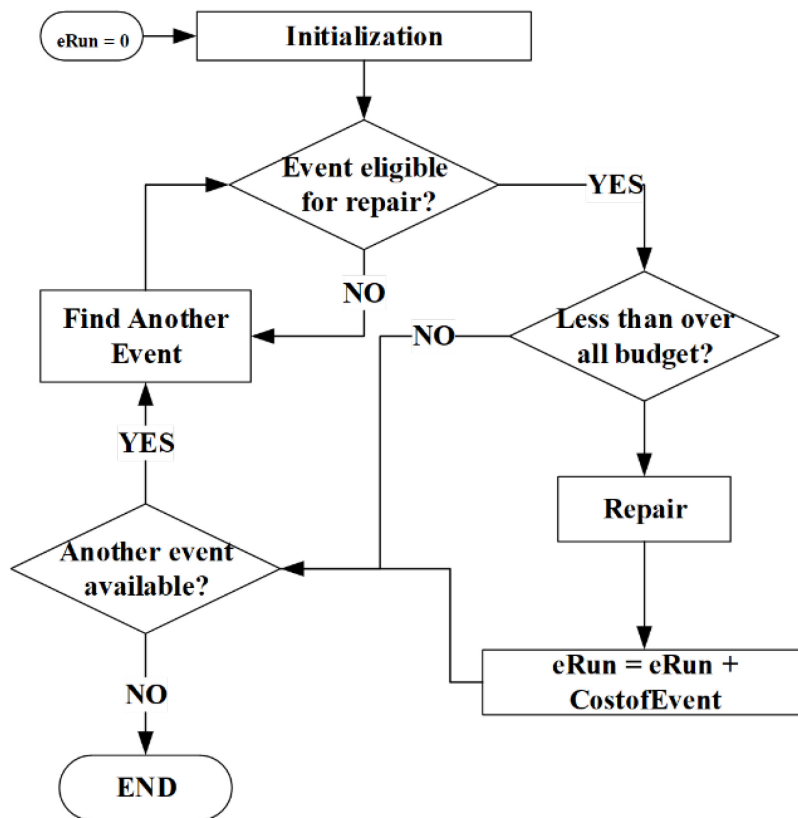


Figure V-12: Updated Budget Equation Methodology

V.1.2 Genetic Algorithm Comparative Model

To compare the Lazy Serpent Algorithm to other algorithms, a genetic algorithm model is developed. The model aims to utilize an approach similar to that of the lazy serpent such that it

uses a 3D environment with motion variables such as costs and time. The purpose of the genetic algorithm model is to identify the order of repair with the best benefit-to-cost ratio.

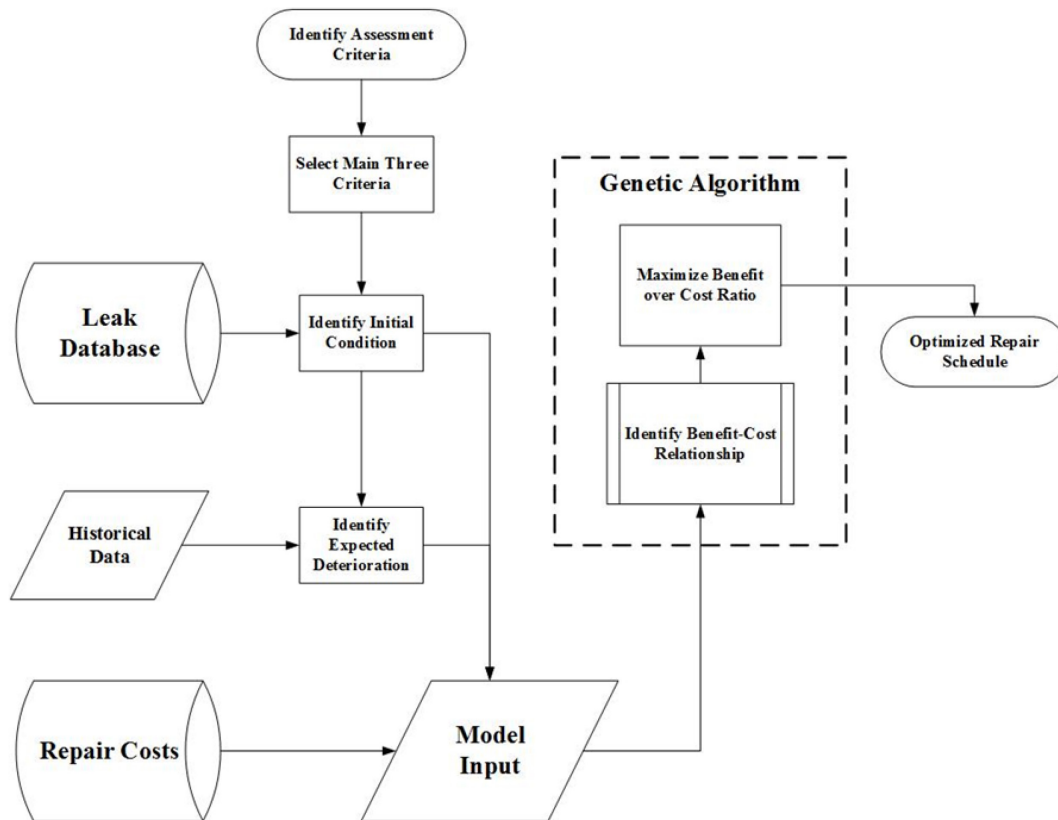


Figure V-13: Genetic Algorithm Model Methodology

Figure V-13 displays the methodology used to develop the genetic algorithm model for the prioritization. The first stage was to identify the main assessment criteria via literature review and common practices. The second stage is to select the three main criteria for assessing each repair event. The next step is to identify the initial condition of each leak and then, based on historical data, to predict the decay and deterioration equations of the leak as a function of time. Furthermore, leak costs and repair costs are identified in a similar manner to that of the lazy serpent to finalize the process that defines the inputs. Before launching the genetic algorithm

model, the benefit-cost relationship amidst the defined criteria should be identified to be maximized. For this model, a fundamental relationship between the cost and the factors is developed and illustrated in Equation (V-16), the equation assuming that a leak is repaired in the best state and what remains is savings and benefits. The factors L, M and N represent the adjustment factors for a specific measure and are used to identify the impact of each factor on the benefit-cost relationship.

$$\frac{B}{C} = \frac{L x_i + M y_i + N z_i}{c_i} \quad (V-16)$$

Where:

- B/C = benefit to cost ratio.
- x, y and z = main assessment factors.
- L, M and N = factor adjustment weights.
- i = represents the number of the event under study.
- c_i = cost of repairing the event i .

The optimization goal can be summarized in Equation (V-17). The goal is to maximize the total value of the benefit to cost ratio such that the total cost (e_{run}) is less than the budget (E) and the total time required to finish the tasks (t_{run}) is less than the allocated time (t). e_{run} is the total cost spent on the completed repairs and t_{run} is the total time used to complete the repairs.

$$\max \sum \frac{B}{C_T} \mid \text{Such that} \begin{cases} e_{run} \leq \text{Budget}(E) \\ t_{run} \leq \text{TimeScheduled}(T) \end{cases} \quad (V-17)$$

V.2 Repair Prioritization Model Implementation

In this section, the models for prioritization are implemented in an example and their results are presented and compared. First, the implementation and results of the GA model are presented and the multiple aspects and results of implementing the Lazy Serpent Algorithm are discussed. The aim is to find the optimal order for repairs, that maximizes the value of the money spent.

V.2.1 Genetic Algorithm Model Implementation and Results

A fictitious sample has been prepared to test the developed genetic algorithm prioritization model. The sample consists of 10 main events that are decaying linearly and have three main assessment criteria extracted from literature (i.e. condition, criticality and consequence of failure). The three factors are presented as place holders to help provide a proof of concept for the developed models. The factors presented require the development of actual models that meet the needs of municipalities. Figure V-14 displays the overall input for constructing the genetic algorithm model. The variable the algorithm tries to optimize is B/C. Furthermore, time requirements and costs are stated for each required repair. For this example, the algorithm is given the time to run all possible scenarios and select the best alternative. Additionally, the algorithm has no constraints regarding time or budget.

Event ID#	Order	Event ID#	Condition	Criticality	COF	Repair period (days)	Repair Cost (\$)	Condition	Criticality	COF	B/C
1	1	1	75.00	50.00	13.88	5	\$ 100.00	75.00	50.00	13.88	13.89
2	3	2	50.00	10.00	74.38	12	\$ 200.00	48.72	10.00	74.38	6.65
3	6	3	65.00	40.00	60.79	9	\$ 400.00	61.56	40.00	60.79	4.06
4	7	4	90.00	30.00	55.80	15	\$ 250.00	85.84	30.00	55.80	6.87
5	10	5	65.00	80.00	46.24	18	\$ 600.00	57.48	80.00	46.24	3.06
6	5	6	40.00	40.00	80.00	10	\$ 348.00	37.36	40.00	80.00	4.52
7	8	7	70.00	30.00	20.00	15	\$ 254.00	64.64	30.00	20.00	4.51
8	4	8	30.00	40.00	50.00	5	\$ 645.00	27.76	40.00	50.00	1.83
9	2	9	70.00	60.00	20.00	11	\$ 200.00	69.60	60.00	20.00	7.48
10	9	10	20.00	80.00	20.00	12	\$ 350.00	13.44	80.00	20.00	3.24
	55										56.11

Figure V-14: Genetic Algorithm Input Excel Sheet

The algorithm requires 68 minutes to arrive at the optimal solution after testing all the possible scenarios. The total number of trials is 28,385, of which 15,456 trials are valid. Figure V-15 displays the improvement path of the solution throughout trials from the initial solution of 55.85 all the way to 56.11 in benefit to cost ratio.

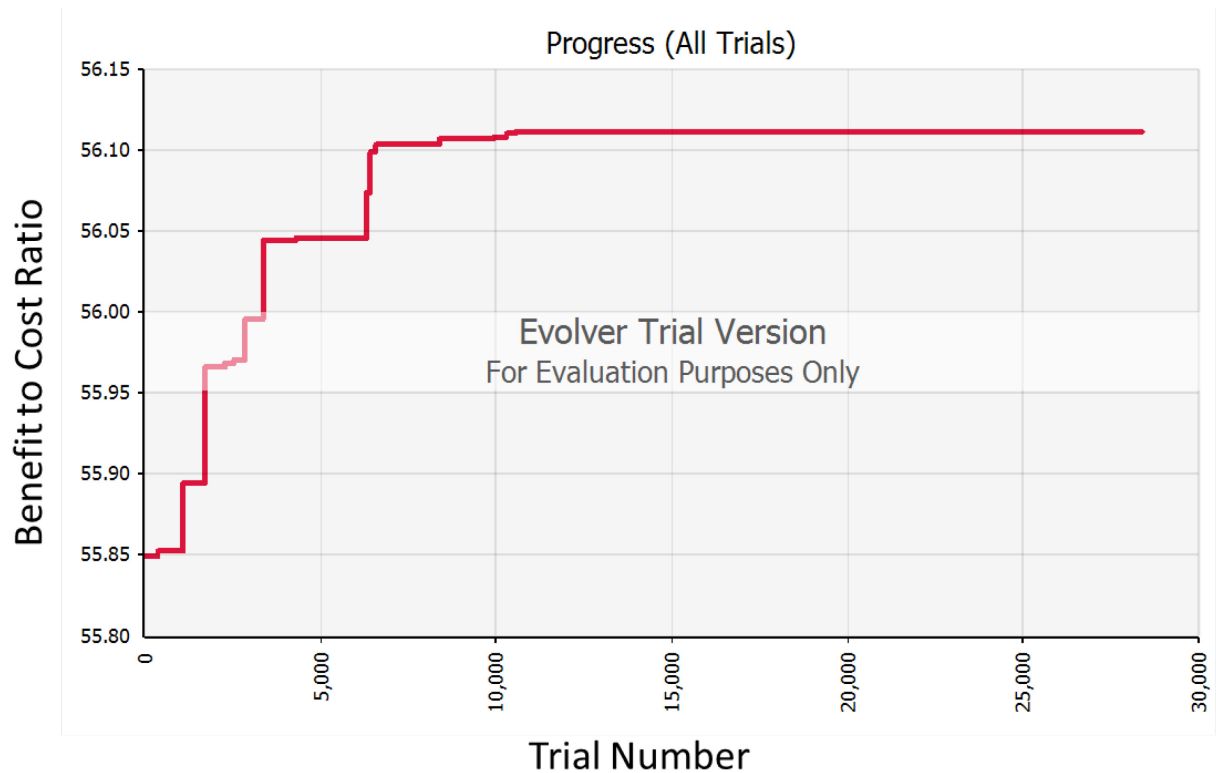


Figure V-15: Genetic Algorithm Progress through Trials

The optimized repair order for the maximized benefit to cost ratio is shown in Figure V-16 where the optimal order of the previous example is displayed as the following order of events: 1, 3, 6, 7, 10, 5, 8, 4, 2 and 9.

	Order									
Event #	1	3	6	7	10	5	8	4	2	9
1	█									
2									█	
3		█								
4								█		
5						█				
6			█							
7				█						
8							█			
9										█
10					█					

Figure V-16: Optimal Repair Order of Fictitious Example

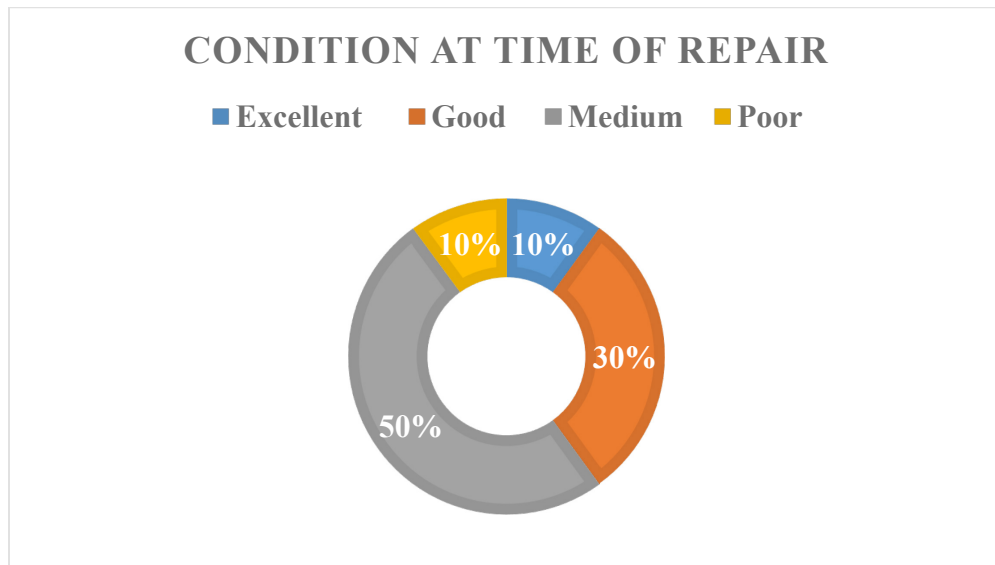


Figure V-17: Distribution of Benefit-Cost Ratio Condition at Repair Time

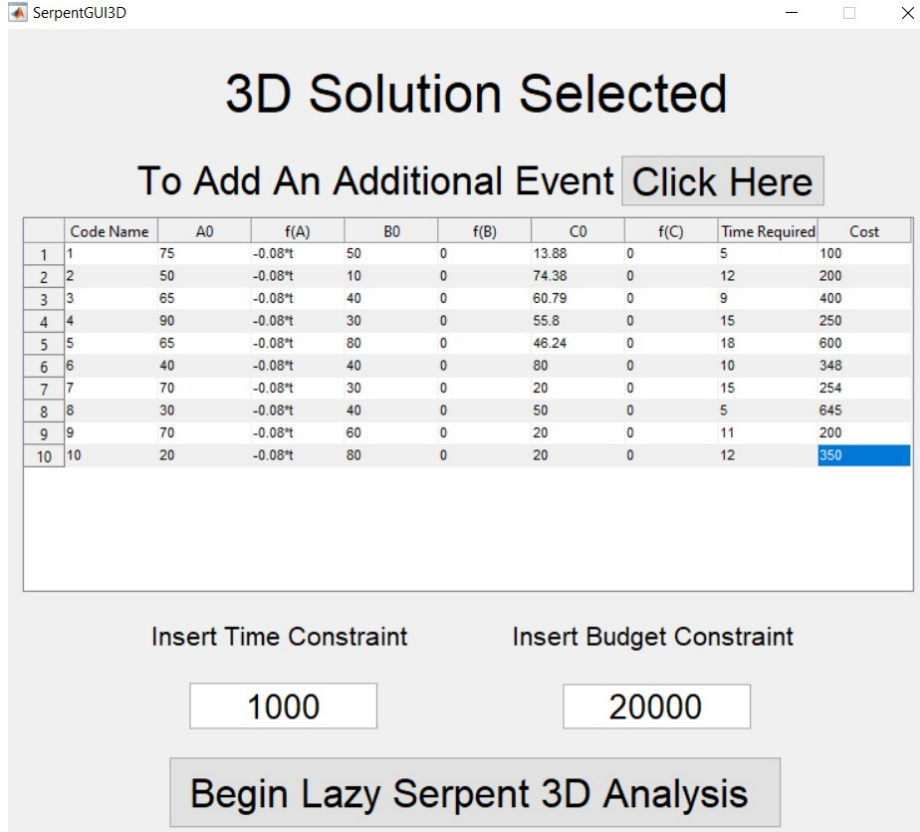
To further assess the quality of the output of the genetic algorithm, the data is categorized into four categories or solution quality levels. The classification is done using the highest possible benefit to cost ratio of a repair event as the maximum possible result, then each event is assessed at repair time against the maximum value. In Figure V-17, a set of ranges is displayed. The ranges are assumed to be as follows: (1) 0% – 15% is a poor-quality solution, (2) 15% – 40% is a

medium quality solution, (3) 40% – 80% is a good quality solution and (4) 80% – 100% is an excellent quality solution. Based on this assessment, the proposed solution includes one solution of excellent quality, three solutions of good quality, five solutions of medium quality and one solution of poor quality. The following values result in an average solution quality of 40.5%, which is a good quality solution based on the proposed assessment.

V.2.2 Lazy Serpent Model Comparative Results

The Lazy Serpent Algorithm is compared to the genetic algorithm using the same sample data in the previous section. Multiple instances of the Lazy Serpent Algorithm are run with different selection criteria definitions. The aim is to explore the aspects of the lazy serpent and see the extent of its power as a prioritization algorithm. Figure V-18 displays the interface of the lazy serpent software developed using MATLAB. The interface consists of multiple parameters, first the codename where the user inputs the name of the event to be prioritized. The codename can be numeric or alphabetic or a mixture of both. The next set of parameters is A0, B0 and C0, which represents the initial condition upon the discovery of the event at the point of assessment. The second set of variables is $f(A)$, $f(B)$ and $f(C)$, which represents the change in the condition of each event. In this example, all events deteriorate in one plane and the deterioration function is $-0.08*t$. Moreover, two variables to identify the required time and cost for each event are added under the tabs Time Required and Cost. Finally, the constraints are added via the Time Constraint and Budget Constraint. The Lazy Serpent Algorithm is run with three different selection criteria: (1) The basic Lazy Serpent Algorithm, which consists of selecting the worst condition event each time and eliminating it, (2) the inverse lazy serpent, which selects the best event each time and (3) the selective lazy serpent, which aims at selecting the most deteriorating

event through projection and if two events have equal deterioration factors, the one with the lowest price will be selected and if the price is the same, the algorithm will move to select the one with the minimum required time. If all factors of selection are equal, an event will be chosen randomly.



3D Solution Selected

To Add An Additional Event [Click Here](#)

	Code	Name	A0	f(A)	B0	f(B)	C0	f(C)	Time Required	Cost
1	1		75	-0.08%	50	0	13.88	0	5	100
2	2		50	-0.08%	10	0	74.38	0	12	200
3	3		65	-0.08%	40	0	60.79	0	9	400
4	4		90	-0.08%	30	0	55.8	0	15	250
5	5		65	-0.08%	80	0	46.24	0	18	600
6	6		40	-0.08%	40	0	80	0	10	348
7	7		70	-0.08%	30	0	20	0	15	254
8	8		30	-0.08%	40	0	50	0	5	645
9	9		70	-0.08%	60	0	20	0	11	200
10	10		20	-0.08%	80	0	20	0	12	350

Insert Time Constraint:

Insert Budget Constraint:

Begin Lazy Serpent 3D Analysis

Figure V-18: Lazy Serpent Software Interface Input

V.2.2.1 The Basic Lazy Serpent Results

The basic lazy serpent selects the closest event to the origin point. So, it selects the most deteriorating events first and goes upwards until it meets a constraint or it finishes all the available events. Figure V-19 shows the results presented by the basic lazy serpent with deterioration preference to be as follows: Starting with event 8, followed by event 7, event 10,

event 2, event 1, event 9, event 3, event 6, event 4 and event 5. The algorithm requires 6 seconds to come up with this solution, as the lazy serpent does not search for all possible outcomes like the genetic algorithm. On the contrary, it goes immediately towards an expected solution. The quality of the solution is measured by the final benefit-to-cost ratio calculated by assuming the remaining three-dimensional values at the point of repair initiation as benefits. The final benefit-to-cost ratio of the proposed solution is calculated to be at 55.81 or 5.581, which is a bit less than that of the first-in-first-out approach.

Event #	Order									
	8	7	10	2	1	9	3	6	4	5
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Figure V-19: Basic Lazy Serpent Repair Priority

V.2.2.2 The Inverse Lazy Serpent Results

The inverse lazy serpent relies on selecting the farthest events from the origin point. In other words, it selects the best events available. The inverse lazy serpent presents a different order from the basic lazy serpent. The results are summarized in Figure V-20. The ordering starts with event 5, followed by the events 4, 6, 3, 9, 2, 1, 10, 7 and event 8. The benefit-to-cost ratio measures the quality of the solution. The ratio is calculated to be 55.31 or 5.531 and the algorithm requires 6 seconds to arrive at the proposed solution.

Event #	Order									
	5	4	6	3	9	2	1	10	7	8
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Figure V-20: Inverse Lazy Serpent Repair Order

V.2.2.3 The Selective Lazy Serpent Results

The selective lazy serpent has a more advanced decision making and selection criteria. The selective lazy serpent aims to select events with a higher deterioration rate than other events, thus it selects the most deteriorating event first. Furthermore, in case deterioration factors are equal between two events, the serpent moves ahead to select the one with the lowest price if the price between two events is equal, the serpent moves to select the one with the lowest time requirements. Finally, if all the aspects are equal, the lazy serpent selects any event from the pool randomly. The results of the selective lazy serpent can be viewed in Figure V-21. The proposed order starts with event 1, followed by event 9, event 2, event 4, event 7, event 6, event 10, event 3, event 5 and event 8. The time required for the development of this solution is 20 seconds and the benefit-to-cost ratio of the solution is calculated to be 56.19 or 5.619.

	Order									
Event #	1	9	2	4	7	6	10	3	5	8
1										
2										
3										
4										
5										
6										
7										
8										
9										
10										

Figure V-21: Selective Lazy Serpent Proposed Repair Order

V.3 Repair Prioritization Results and Conclusions

All the prioritization results previously mentioned are summarized in Table V-1. Table V-1 compares and analyzes the impact of each respective approach. The approaches and algorithms under study are (1) Inverse Lazy Serpent, (2) Basic Lazy Serpent, (3) First In First Out (FIFO), (4) Genetic Algorithm and (5) Selective Lazy Serpent. The initial benefit-to-cost value of the FIFO approach is calculated, and it is 5.585 value per dollar. Although the results are comparatively good, the FIFO approach cannot be relied on to prioritize repair events regularly, as it is random and unpredictable and thus its results can be good or bad with no consistency. The inverse lazy serpent gives the least benefit-to-cost ratio and therefore it can be concluded that leaving the worst for last can be a wrong approach and costlier than the remaining approaches. The basic lazy serpent has more acceptable results with a value of 5.581, which is close to the original 5.585 presented by the FIFO approach. Both the inverse lazy serpent and the basic lazy serpent require 8 seconds to give their results, which is considerably fast. The genetic algorithm presented the first improvement to the order of repairs with a benefit-to-cost ratio of 5.611, which is higher than those of the FIFO approach, the inverse lazy serpent and the basic

lazy serpent. The genetic algorithm requires 67 minutes to solve because the genetic algorithm tries to search throughout all the possibilities to arrive at a feasible solution. Finally, the selective lazy serpent gives a solution with the highest benefit-to-cost ratio with a value of 5.619 or 56.19. The selective lazy serpent requires 20 seconds to develop and present the priority of repairs. The main reason for the selective lazy serpent to beat the genetic algorithm in terms of computational time is that the selective lazy serpent goes immediately towards building a solution in contrast to scouring multiple options as the genetic algorithm.

Table V-1: Comparison of Prioritization Models

Algorithm / Approach	Inverse Lazy Serpent	Basic Lazy Serpent	FIFO	Genetic Algorithm	Selective Lazy Serpent
Results (B/C ratio)	5.531	5.581	5.585	5.611	5.619
Time Consumed	8 seconds	8 seconds	0	67 minutes	20 seconds

CHAPTER VI: DEVELOPED AUTOMATED TOOLS

The fourth objective of this thesis is to automate the developed models. Therefore, multiple software tools are developed in this chapter for the automatic analysis of the data that is provided by the developed models. Specifically, three tools are developed: (1) Acoustic Leak Detection tool, (2) Acoustic Leak Pinpointing Tool and (3) Automated Lazy Serpent Tool.

VI.1 Acoustic Leak Detection Tool

The platform used by the water services agency in the city of Montreal suffers has one critical issue and that is false alarms. False alarms can be in the form of identified leaks or suspected leaks. This created the need for a tool that can distinguish between the sounds of leak origin and non-leak origin. Therefore, an automated tool is developed using MATLAB for the analysis of the signals collected by the noise loggers that spread in the downtown of the city of Montreal.

VI.1.1 Acoustic Leak Detection Tool Design

The architecture of the developed tool can be described by Figure VI-1. After launching the tool, a user interface is displayed. The user interface needs the wav file selection button to be clicked, or the interface will remain idle. Once the wav file selection button is clicked, a new interface will open. The new interface allows the users to browse through their files and folders to select the one that requires analysis and assessment. After selecting the file, the tool checks if the file is in wav format. If the file is of a different format, an error would be displayed and the software will exit. If the file is correct, the tool will move ahead, and conduct Fourier transform of the signal. Based on the Fourier transform analysis, three main indicators are collected and the shape

of the signal is displayed. The three indicators are (1) FI_{50Hz}, (2) Level and (3) Spread of the signal. FI_{50Hz} represents the number of waves within the sound signal, that carry a frequency of 50Hz or less. Level represents the maximum sound level reached by the signal in decibels (dB). Spread represents the difference between the highest measured sound (level) and the lowest measured sound. Each factor will be analyzed against a previously developed threshold. If the value of the indicator exceeds the threshold, that indicator identifies that the sound is not generated by a leak. If the indicator exceeds the value of the threshold, a value of 1 is added to the aggregator. Otherwise, a value of zero is added. After all, indicators are tested against their respective thresholds and the value of the aggregator is assessed. If the value of the aggregator is greater than 1, the sound file is not of leak origin. Otherwise, the sound would be of leak origin if the aggregator has a value of 1 or 0. The tool moves onwards to display the result of the conducted assessment by stating “Leak Sound” or “Non-Leak Sound.”

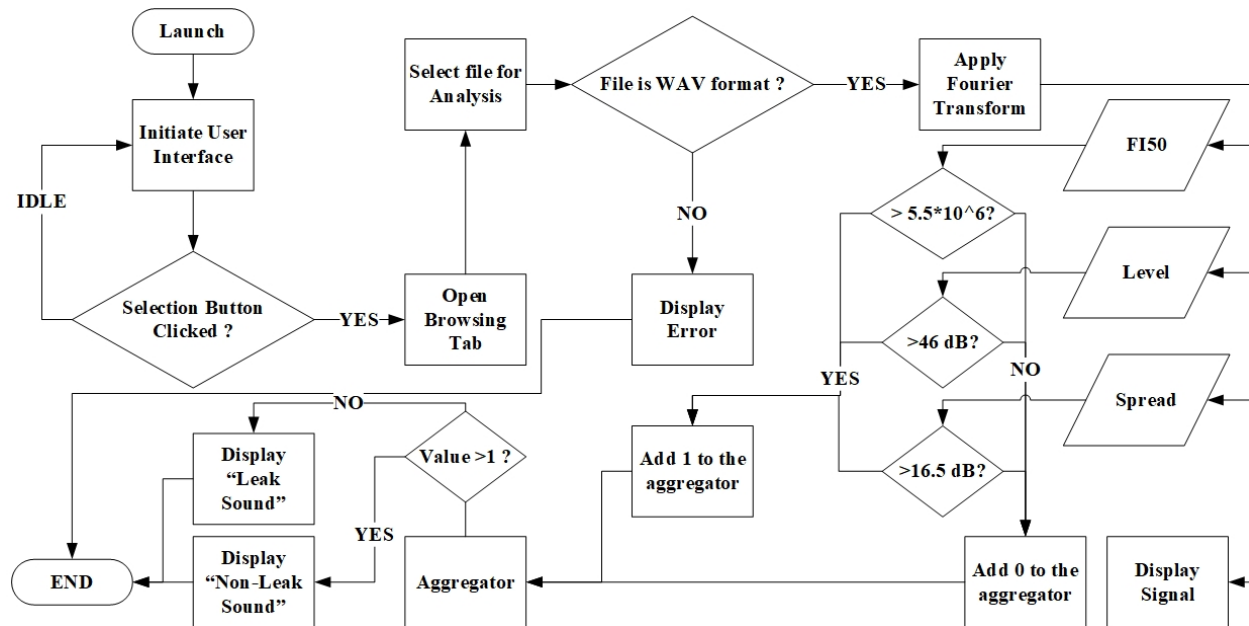


Figure VI-1: Leak Detection Tool Processing Flowchart

VI.1.2 Acoustic Leak Detection Tool Manual

The acoustic leak identifier is designed based on the previously developed research to help water services agents in the city of Montreal distinguish between pump sounds from leak sounds. Both sounds categories are received by the online software implemented for the city of Montreal by Guttermann, a leak detection company. The following steps and images describe how the software is used. It is used after the mp3 sound file of a suspected leak of pump sound is downloaded. The software is beneficial for identifying the origin of the sound without the need for on-site visits.

Step 1: Launch Software

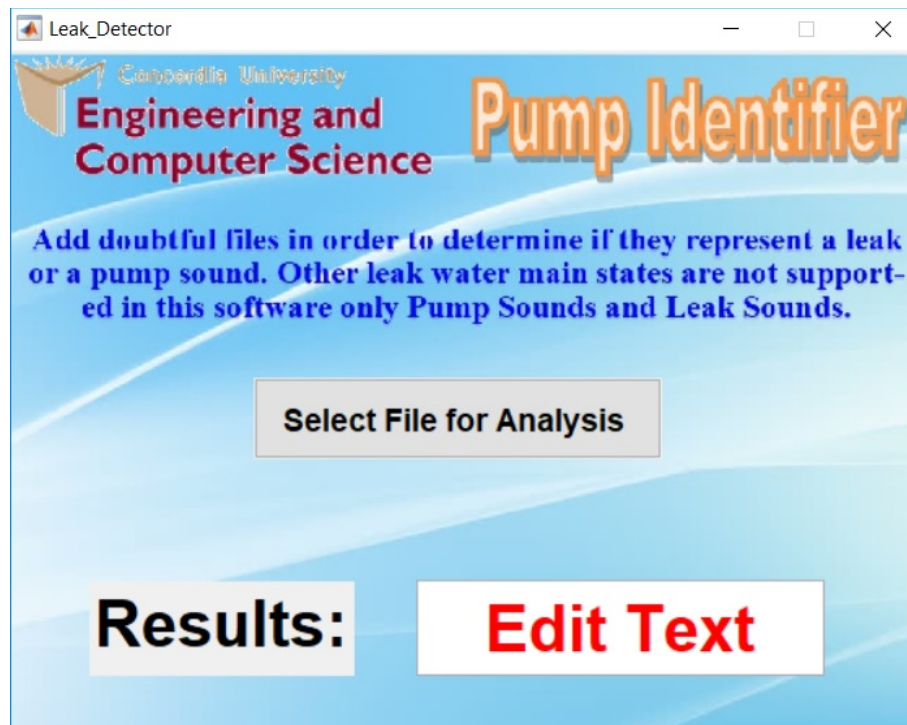


Figure VI-2: Pump and Leak Identifier Interface

The initial step is to launch the execution file of the software. Once double clicked, the interface in Figure VI-2 will show up. The interface is comprised of a paragraph of instructions and the name of the software and the logo of the engineering and computer science department at Concordia University. The interface has one central button and it is dubbed “Select File for Analysis”. To begin the analysis, click on this button and a new window appears to help the user select the desired file to be analyzed. The interface also holds a results section with an “Edit Text” area marked in red.

Step 2: Select Files for Analysis

An interface will open to allow the user to browse for the wav file as in Figure VI-3. The user can select the suspected file and the file analysis begins in the software.

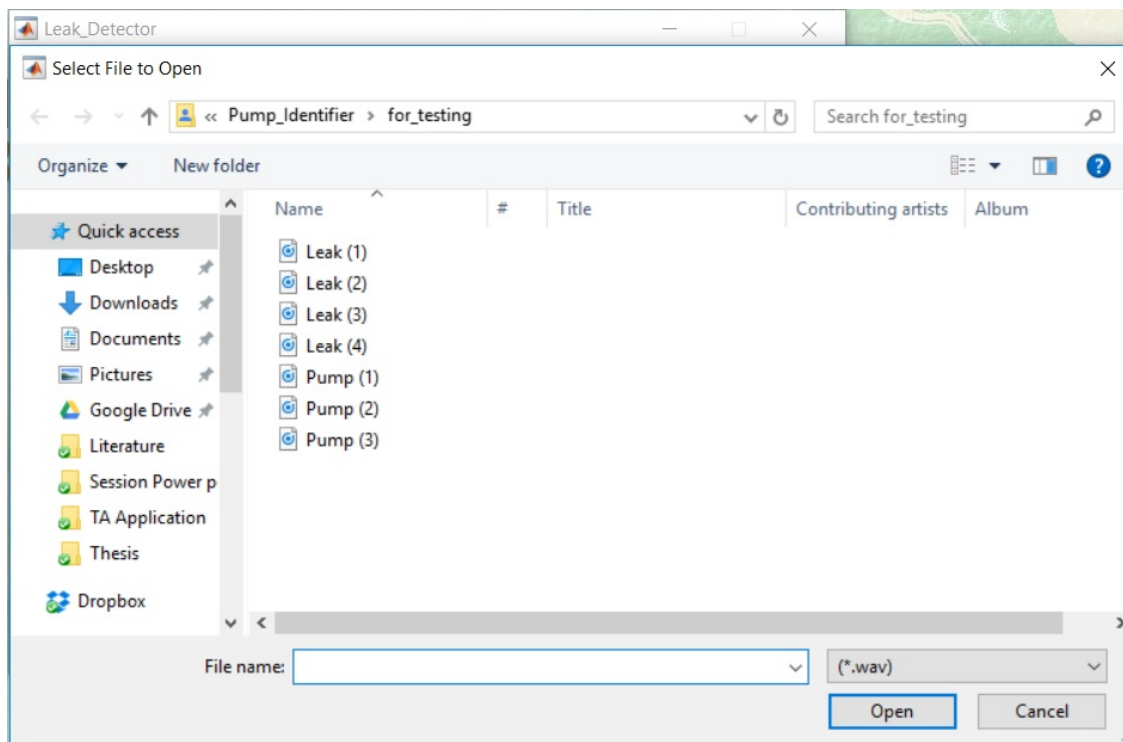


Figure VI-3: Software Browsing Interface

Step 3: Signal Shape Diagram

The software first moves towards analyzing the characteristics of the signal and presenting its shape, as in Figure VI-4. The figure shows the sound levels of the signal against its time component and how the sound carries through time to allow checking for a typical pattern and identifying any possible anomalies within the signal. The image also presents a typical signal pattern, which shows the final result of the analysis by the software.

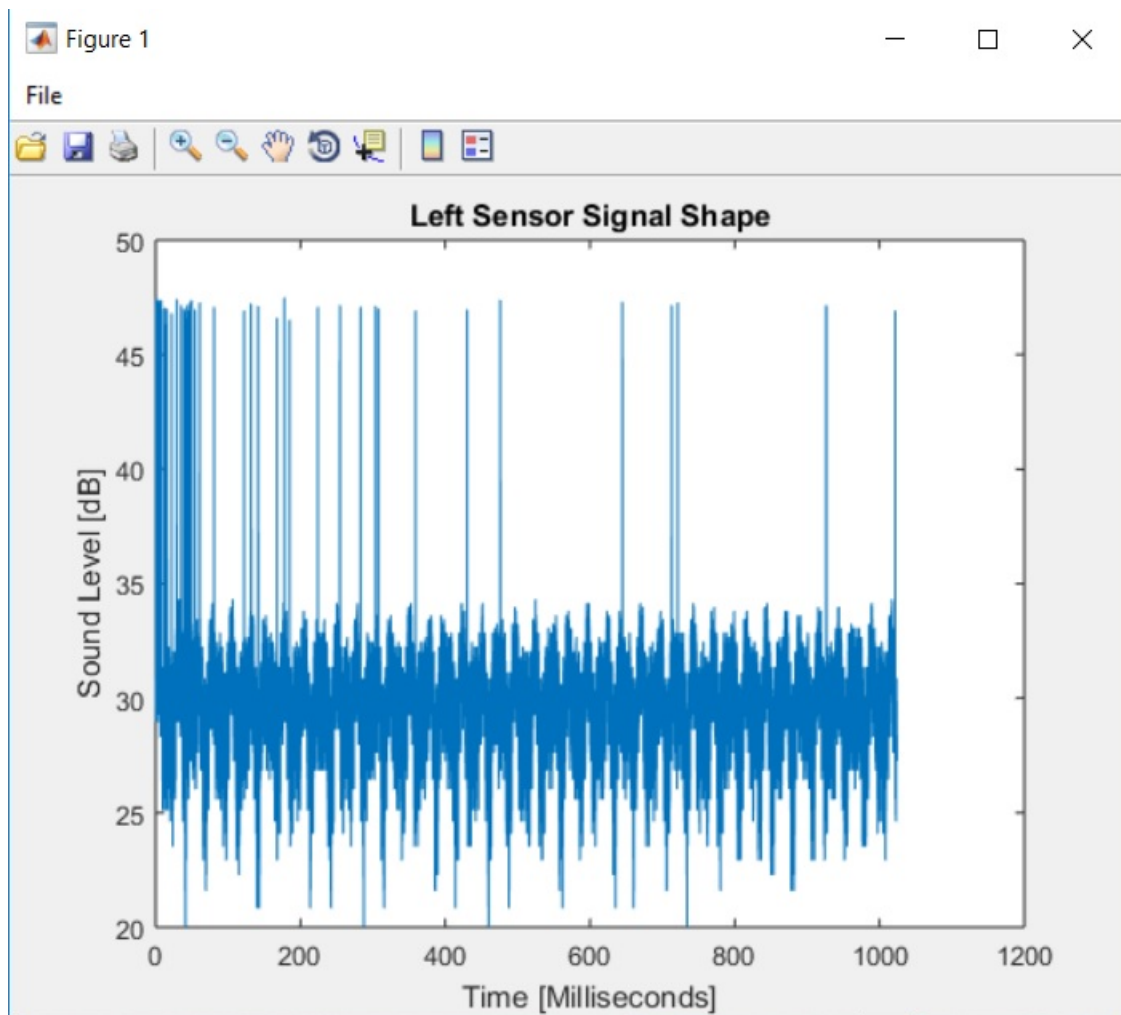


Figure VI-4: Wave Representation of an Analyzed File

Step 4: Analysis Results and State Identification

Finally, the software displays the result for the analyzed wav file in editable text file presented next to the word “Results.” The possible outputs are either “Pump Sound” as in Figure VI-5 or “Leak Sound.” When the result indicates a pump sound, the software identifies the wav file to be on non-leak origin. Otherwise, the wav file is that of a leak.



Figure VI-5: Results of the Analysis

VI.2 Acoustic Leak Pinpointing Tool

Improvements in leak pinpointing for the acoustic system installed in downtown Montreal is another aspect of the work conducted with the water services agency of the city of Montreal. As

part of this research, a tool is developed to automate the signals analysis that indicate a leak state and then use the results of the analysis to deduct the location of the leak within the network between the two sensors detecting the leak.

VI.2.1 Acoustic Leak Pinpointing Tool Design

The leak pinpointing tool design can be further illustrated in Figure VI-6. After initializing the tool, an interface pops up that contains an area for distance input and a begin initialization button. Once the button is clicked, a browsing interface pops up asking for the selection of the leak sound files in wav format. The user has to select the left sensor first and then the right sensor for the software to operate properly. If the sound files are any other format, an error message is displayed and the software exits. If the signals are in the proper format, the sound files will be analyzed and their data is collected, subject to the previously developed regression equations. Afterwards, an output screen is displayed showing the distances X_L and X_R from the left and right sensors towards the leak. The user now has the option of closing the software.

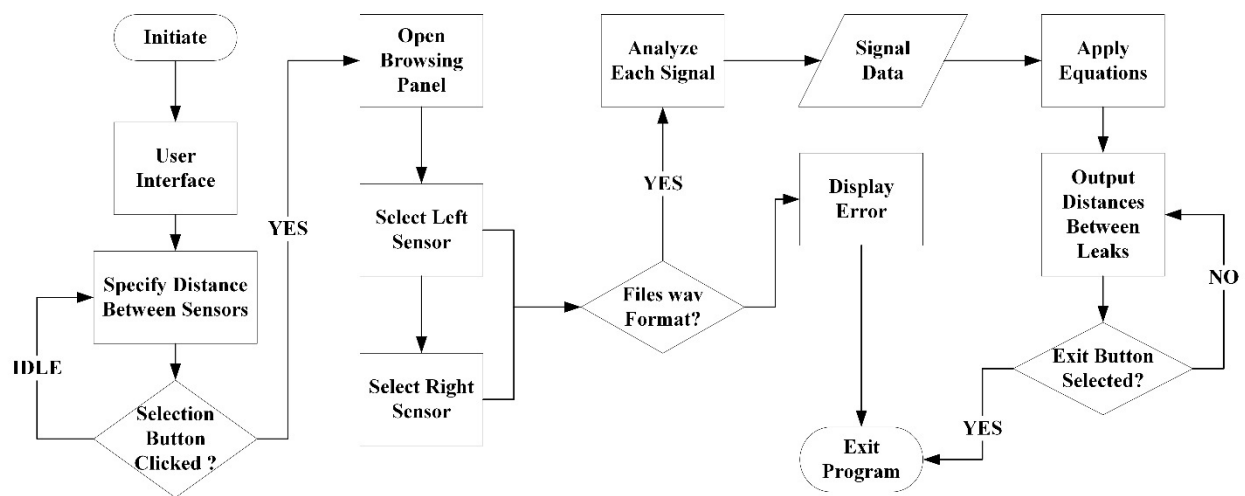


Figure VI-6: Leak Pinpointing Tool Design Flowchart

VI.2.2 Acoustic Leak Pinpointing Tool Manual

This section guides the user of the pinpointing tool along the steps required to make the tool operate accurately and efficiently. The operation is comprised of five main steps: Initialization, distance indication, file selection, signal analysis and distance output. The tool takes two leak signals surrounding a confirmed leak and utilizes the sound data to present an accurate estimate of the leak location based on the previously developed models. The tool complements the Guttermann software used by the water services agency in the city of Montreal and uses leak sounds detected by the listeners in the automated leak detection system.

Step 1: Tool Initialization and Tool Interface

After the executable file is initiated, the interface appears on the screen, as illustrated by Figure VI-7. The interface contains tips and steps for using the software and an explanation of how to interpret the results. It also contains an editable area for inserting the distance and a push button named “Pinpoint Leak.”

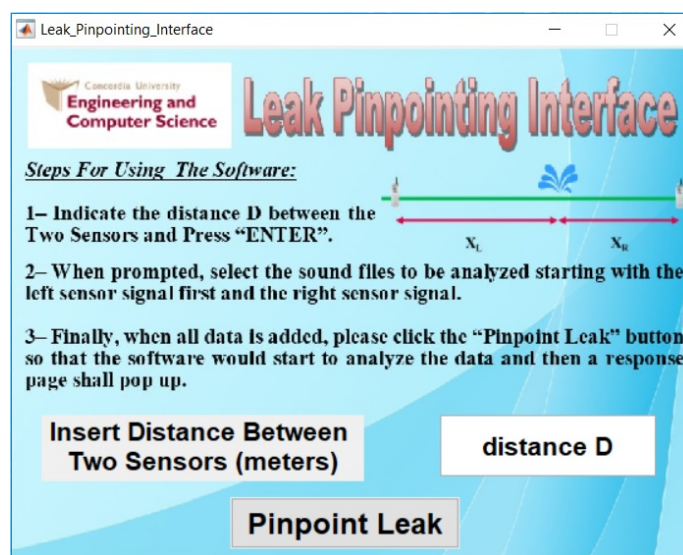


Figure VI-7: Initial Leak Pinpointing Interface

Step 2: Distance Input

The next step is to identify the distance. If the distance is not identified, the software will face an error while running. The distance is inserted in the allocated space labeled “distance D”. The value of the distance is in meters and can be either in integer format or decimal format. For example, in Figure VI-8, the value of the distance is 254 meters. After writing the distance in the allocated area, it is necessary to press the “Enter” button for the tool to read the value internally. After inserting the distance and properly assigning it to the tool, the user can proceed with the analysis through clicking the pinpoint leak push button that is encircled in red in Figure VI-8.

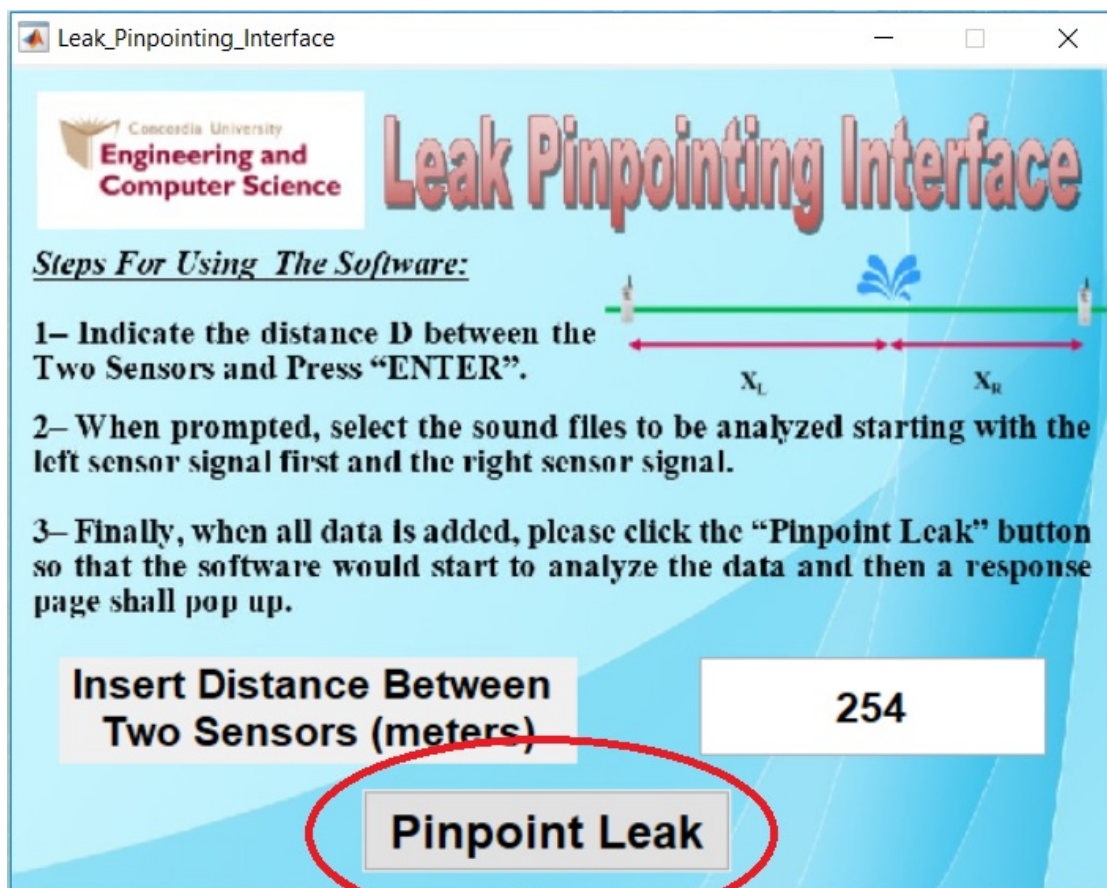


Figure VI-8: Adding The Value of Distance D

Step 3: File Selection

After clicking the initialization button, a browsing page appears as displayed in Figure VI-9. The pane allows the user to navigate to the location where the sound files are stored. The user must start by selecting the left sound signal first and then press “Open”. Afterwards, the navigating pane pops up again to prompt the user to select the sound signal on the right end of the leak to conduct the analysis. Once both files are selected, the tool proceeds towards analyzing the sound files and applying the regression equations to identify the location of the leak.

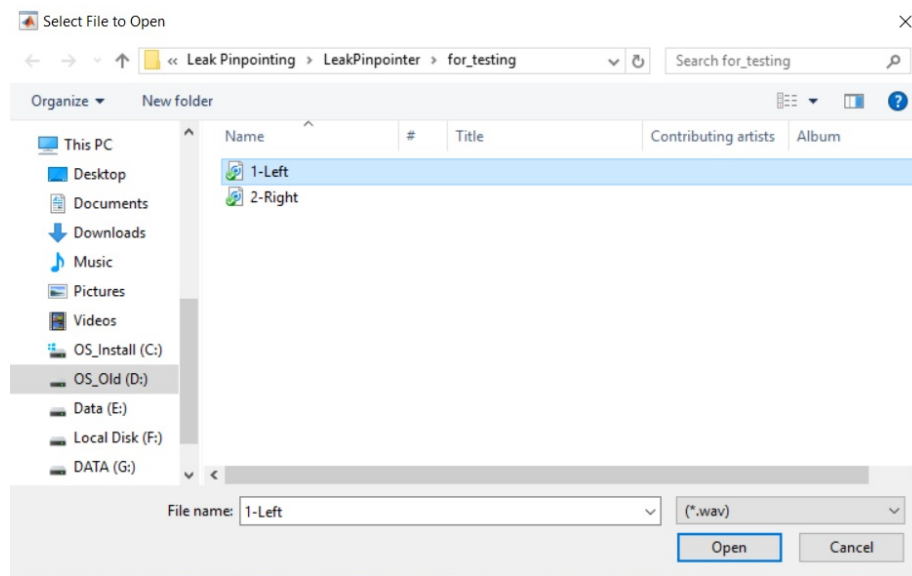


Figure VI-9: Pinpointing Tool Browsing Panel

Step 4: Signal Analysis Output

The analysis of the left and right signals is displayed by the software to allow the user to identify any sound anomalies within the signal. Figure VI-10 shows that the tool outputs two drawn signals for each sound file with the sound level (dB) against time (milliseconds).

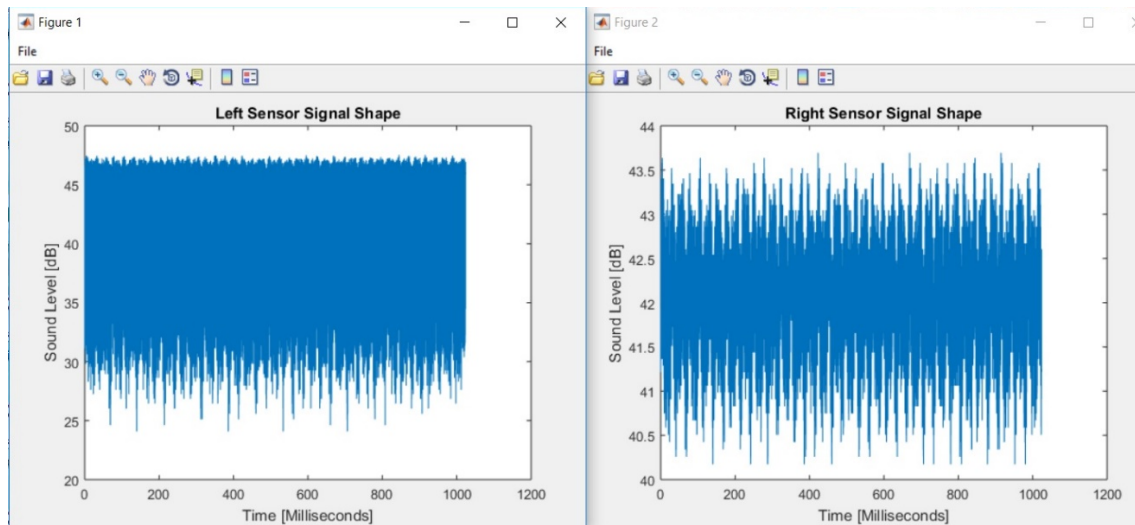


Figure VI-10: Pinpointing Signal Analysis Results

Step 5: Predicted Distances Output

When the analysis concludes, a new interface appears showing the results of X_L , X_R , PX_L and PX_R . X_L and X_R are in meters and represent, as illustrated in Figure VI-11, the distance from the left sensor to the leak and the distance from the right sensor to the leak, respectively. Meanwhile, PX_L and PX_R represent the percentage of the total distance from the left sensor to the leak and from the right sensor to the leak, respectively.

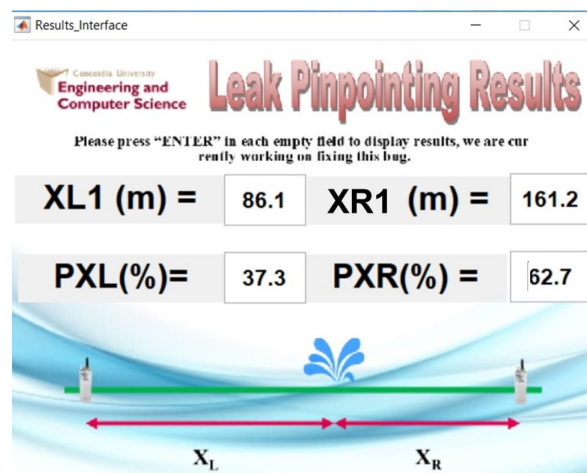


Figure VI-11: Leak Pinpointing Result Interface

VI.3 Automated Lazy Serpent Tool

To apply the Lazy Serpent Algorithm to a substantial quantity of events, a computer-based tool is required to automate the process. So, the automated lazy serpent software is developed for the automatic and fast analysis of 2D and 3D events in the analysis space. This section provides an overview of the tool design and a user manual to explain how the developed tool operates.

VI.3.1 Automated Lazy Serpent Tool Design

The design of the tool can be summarized by the flowchart in Figure VI-12. The first step is to initialize the tool. Then, a user interface for data input appears, to allow the user to input the data manually. The user interface includes areas for adding constraints and for adding extra rows for the analysis. After adding the input, the user can initialize the software by clicking the begin analysis button. The analysis starts by organizing the data in a matrix format and creating an event dataset. The analysis starts by organizing the data in a matrix format and creating an event dataset.

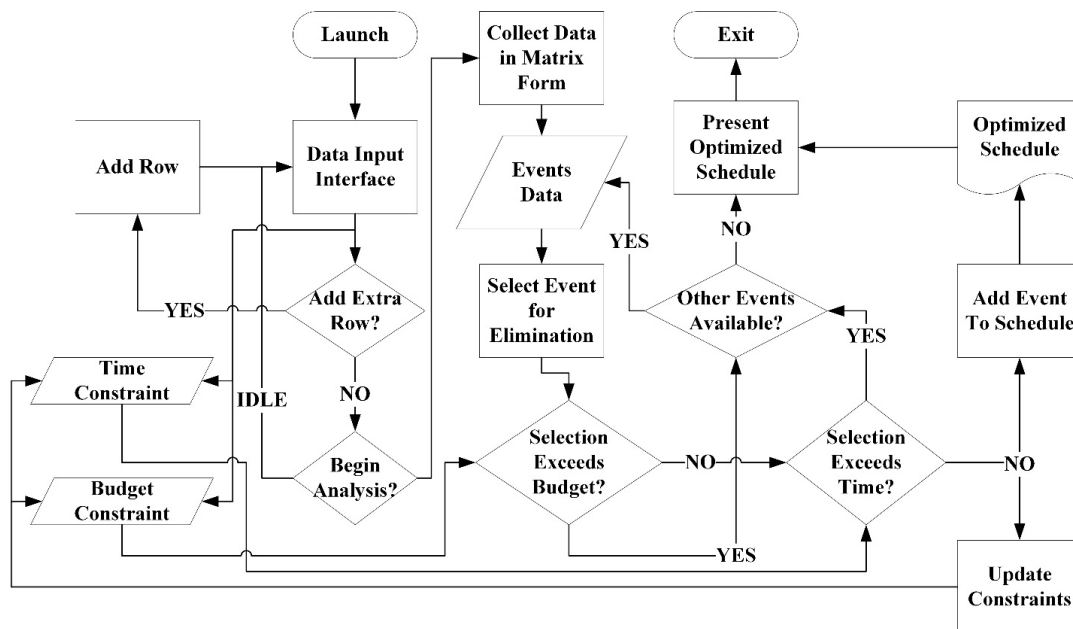


Figure VI-12: Automated Lazy Serpent Design Flowchart

Based on the predefined serpent selection criteria, the serpent moves towards selecting the best event to initiate the prioritization process. Each selected event is examined to realize if it meets the time constraints and the cost constraints. After testing, the event is eliminated from the initial dataset. If the event has met the two constraints, it is added to the optimized repair order. The tool continues looping in this manner until all events are analyzed. Finally, the tool presents the optimized repair order when all events are analyzed.

VI.3.2 Automated Lazy Serpent Tool Manual

This section presents the steps required to operate the automated lazy serpent tool properly. The operation of the tool follows four main steps starting with mode selection, followed by identifying constraints, data input and finally analysis and results.

Step 1: Select Lazy Serpent Mode

In the initial screen illustrated in Figure VI-13, the user can select the operating mode between two-dimensional and three-dimensional, based on the representation of the events.

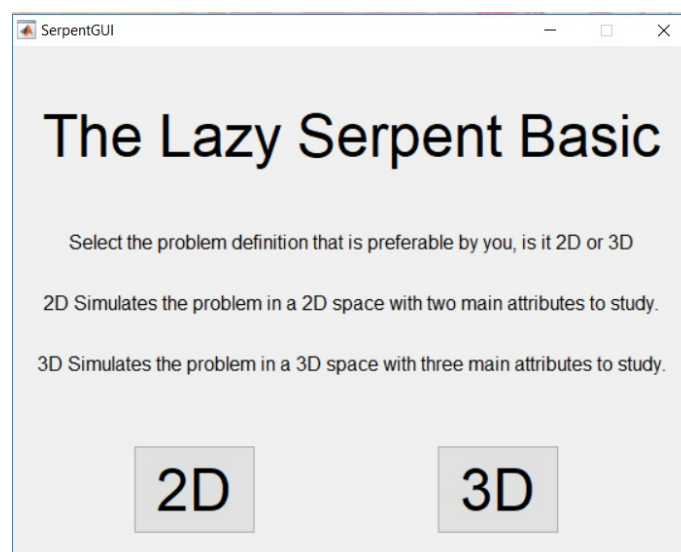


Figure VI-13: Automated Lazy Serpent Initial Interface

Step 2: Defining Constraints

After selecting the mode, the respective interface of the selected mode appears. In this example, the two-dimensional analysis is selected and, accordingly, Figure VI-14 appears on the screen. The user can add extra rows beyond the four default rows by using the “Click Here” button encircled in blue. Furthermore, the user can type the constraints in the specified editable text locations encircled in red and then press “Enter.” If the user wishes to identify an infinite number for a constraint, the user must insert a number greater than any possible combination and press “Enter.”

SerpentGUI2D

2D Solution Selected

To Add An Additional Event [Click Here](#)

	Code Name	A0	f(A)	B0	f(B)	Time Required	Cost
1							
2							
3							
4							

Insert Time Constraint

100

Insert Budget Constraint

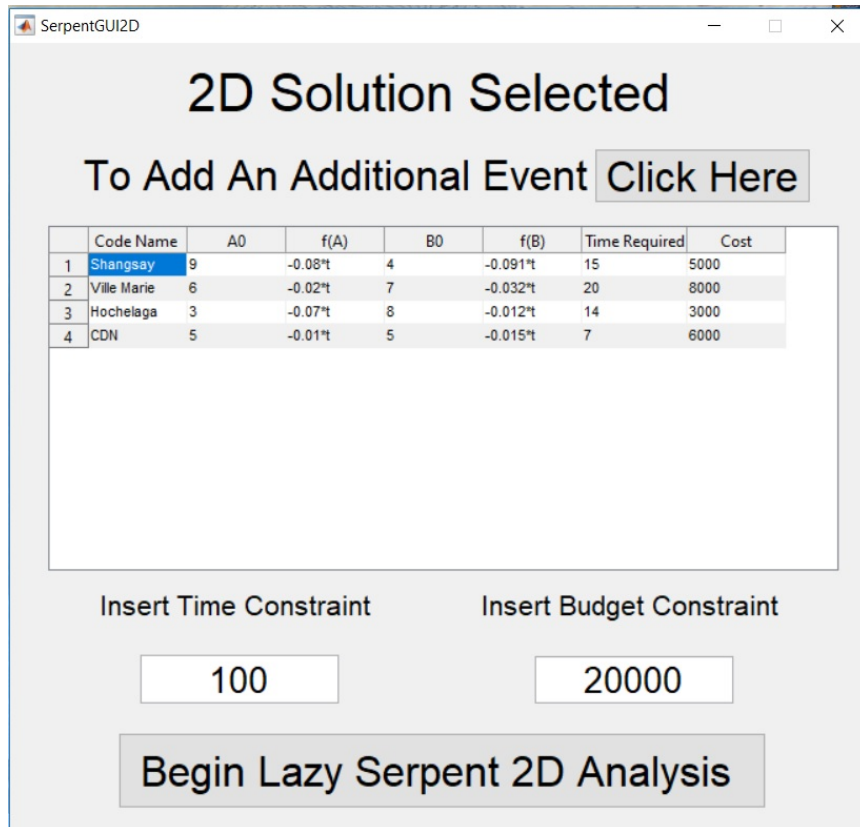
20000

Begin Lazy Serpent 2D Analysis

Figure VI-14: Automated Lazy Serpent Constraint Definition

Step 3: Input Data

The solution input field is comprised of five main input areas. The first area is the code name where the representing code of the event can be inserted. The code name can be a number or a name or even a mix of both. The second is the initial state values area. The example shown in Figure VI-15 shows A0 column and B0 column where the initial states of each event are inserted as integers. The third area is the change equations. In the figure, columns f(A) and f(B) represent the equations of change through time (t). t must always be used to represent the time variable. The equations must always comply with the MATLAB rules and definitions for equation writing. The fourth and fifth areas represent the time required to tackle the event and the funds required for the task respectively.



The screenshot shows the SerpentGUI2D window. At the top, it says "2D Solution Selected" and "To Add An Additional Event Click Here". Below this is a table with 7 columns: Code Name, A0, f(A), B0, f(B), Time Required, and Cost. The table contains 4 rows of data. Below the table are two input fields: "Insert Time Constraint" with the value 100, and "Insert Budget Constraint" with the value 20000. At the bottom is a large button labeled "Begin Lazy Serpent 2D Analysis".

	Code Name	A0	f(A)	B0	f(B)	Time Required	Cost
1	Shangsay	9	-0.08*t	4	-0.091*t	15	5000
2	Ville Marie	6	-0.02*t	7	-0.032*t	20	8000
3	Hochelaga	3	-0.07*t	8	-0.012*t	14	3000
4	CDN	5	-0.01*t	5	-0.015*t	7	6000

Insert Time Constraint: 100

Insert Budget Constraint: 20000

Begin Lazy Serpent 2D Analysis

Figure VI-15: Automated Lazy Serpent Data Input

Step 4: Solution Output

Finally, after pressing the “Begin Lazy Serpent 2D Analysis” button, the tool automatically starts to calculate its path towards a solution that complies with the predefined constraints and event search criteria. Once the algorithm meets one of the constraints or the event database is cleared, the algorithm halts and the output is displayed as illustrated by Figure VI-16. The figure shows that the proposed optimal repair order to meet the required constraints is to repair CDN first, followed by Hochelaga and finally Shangsay. Ville Marie is not in the list as the event is the least critical based on the predefined criteria and it exceeds one or two of the constraints.

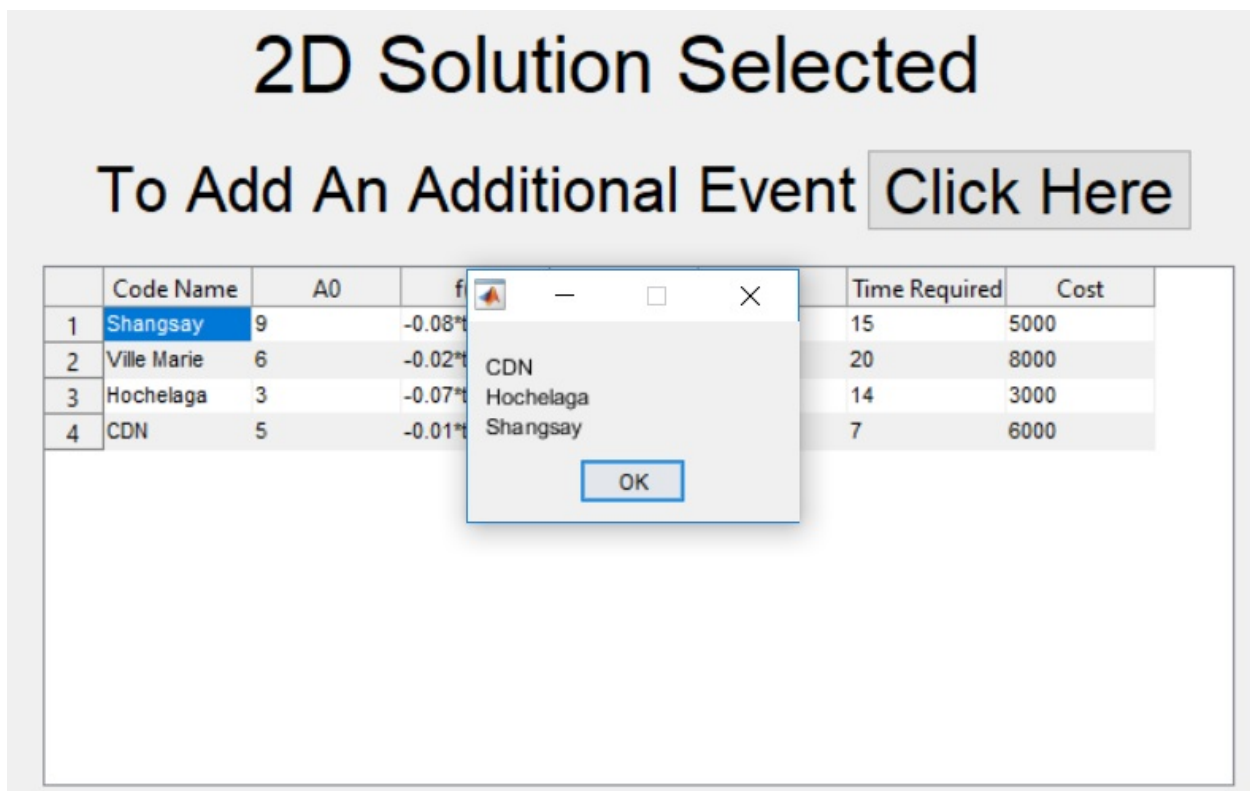


Figure VI-16: Automated Lazy Serpent Result Output

CHAPTER VII: RESEARCH CONTRIBUTIONS AND FUTURE WORK

VII.1 Summary and Conclusions

In conclusion, accelerometers and noise loggers are powerful devices capable of detecting and pinpointing leaks in the water network and both technologies can help the move towards static leak detection systems using internet of things. After vibration signals are analyzed, their outcome (MIE) can be used via classification techniques for identifying the existence of a leak and its size. MIE proves to be an effective parameter for assessing vibration signals and identifying the existence of leaks through an MIE threshold. When the data is coupled with regression analysis, the leak location can be predicted, via accelerometers, with an accuracy of 85% and a range of 25 centimeters. Accuracy can raise to 100%. Accelerometers have proved to be robust devices to create a capable static leak detection and pinpointing system without the aid of other leak detection technologies. Meanwhile, further research is required on their behavior in various environments and buried pipelines. Such research can include multiple event testing and signal response to filtering.

The hypersensitivity of accelerometers can help identify and locate small leaks if adequately filtered. Otherwise, the readings may contain interfering signals that can trigger false alarms or divert leak pinpointing. Similarly, for acoustic noise loggers, the acoustic signal is analyzed and transformed into a series of variables, mainly level, spread and frequency. The variables are utilized to identify the existence of leaks when supported by classification techniques. The technology can also aid in identifying whether the recorded sound is a leak or an interfering sound, using classification techniques coupled with an aggregation mechanism. The accuracy of such models can reach up to 100%. As for leak pinpointing, regression analysis is used with collected acoustic data to determine the exact location of a suspected leak. Acoustic listeners

tend to be a more popular static leak detection technology due to the simplicity of analyzing and filtering acoustic signals and the vast knowledge of their operation, making their results in leak detection much more accurate than accelerometers in an actual field situation. Accelerometers will present an optimistic future in leak detection if sufficient research and more extensive experiments are conducted while using them. Those two modes of detection can provide cities with the capability to save money through early detection if used alone or together for a more powerful detection system.

Furthermore, leak repair prioritization can be performed using two approaches: (1) Genetic Algorithm and (2) the Lazy Serpent Algorithm. Both approaches have proved to provide a better solution than the usual first in first out (FIFO) approach. The genetic algorithm can provide a good solution but, when it comes to the time requirement of GA, the time to propose a solution to a small problem is substantially large, and accordingly, bigger scale projects may not attain a good solution quickly enough. On the other hand, the Lazy Serpent Algorithm can provide better solutions in some cases, yet in much lesser time as the Lazy Serpent Algorithm goes directly towards finding a solution instead of searching multiple possibilities and progressing. Through its search mechanism, the lazy serpent starts ordering events one by one as they are being simulated through time. Accordingly, municipalities now can have more directed leak repair prioritization orders that fit their criteria and maximize return on their expenditure and save money. Therefore, combining simulation and prediction strategies with optimization can provide municipalities with a sense of visualization allowing them to see how leaks behave in the predictable future. The strategic prioritization of such independent repair events is a key factor in saving money for governmental organizations as they allow the proper prediction and

implementation of future scenarios. The use of prediction models with the inverse pyramid solution is a powerful tool in this regard.

VII.2 Research Contributions

This research aims at solving the leak detection problem for municipal bodies by providing an in-house real-time leak detection monitoring system, besides a repair prioritization algorithm. On the level of accelerometers, the final output can provide immediate indications of the formation of a leak and pinpoint the leak with a 25-cm radius accuracy. On the level of loggers, the model presented a high accuracy of pinpointing by identifying the exact leak location with a 20-cm accuracy. The question beyond leak detection is that of leak repair prioritization. Using the developed Lazy Serpent Algorithm, municipalities will be able to identify the leaks that account the most critical, based on their criteria, simulate the progression of those leaks to a certain extent and identify the optimal repair priority.

Therefore, the primary contributions of this research are summed up as follows:

- 1- A comprehensive literature review is performed, shedding light on the current state of the art in leak detection, and the most suitable devices for this research is selected.
- 2- A set of experiments and models are conducted to help enhance the accuracy and the detection capabilities of static leak detection systems that utilize noise loggers and accelerometers.
- 3- Basic experiments that utilize accelerometers in leak detection networks are done. The experiments show that accelerometers have potential in the field of leak detection with extra experimentation and research.

- 4- A model is developed, capable of identifying the existence of leaks in ductile iron as well as PVC pipelines using accelerometers.
- 5- A model is developed, capable of identifying the size of a leak in ductile iron as well as PVC pipelines using accelerometers.
- 6- A model is developed, capable of pinpointing a leak in ductile iron as well as PVC pipelines using accelerometers.
- 7- Possible approaches to assess the decay of vibration signals in ductile iron and PVC pipelines are analyzed.
- 8- A model is developed, capable of identifying the existence of leaks using acoustic noise loggers for the network surveillance project in the city of Montreal.
- 9- A model is developed, capable of differentiating between the sound of leaks and the sound of pumps within the water network for the pilot project for water network surveillance in the city of Montreal.
- 10- A model is developed, capable of pinpointing the leak within a network monitored by acoustic noise loggers for the water network surveillance project in the city of Montreal.
- 11- A genetic algorithm model that aims to optimize repair ordering is developed.
- 12- An algorithm is developed, capable of identifying the best possible repair ordering for identified leaks in a small time-frame.
- 13- An aggregation technique is developed, that helps improve binary classification models through using their collective thinking.

VII.3 Research Limitations

Like in any research, there are several limitations in this particular work and due to the broadness of its scope, the limitations are broad. They can be summarized as follows:

- 1- The designed experiments for accelerometers study the necessary conditions for operation. The conditions are the existence of a pressurized pipelines and induced leaks. In reality, multiple other factors exist like soil type, branching, and leak progress which should be included in future experimentation. Exploring the actual operational conditions allows the comprehension of the requirements for using accelerometers in the real world.
- 2- Lack of filtering for both vibration signals and acoustic signals: The utilization of proper filtering techniques is assumed to have an impact on the quality of the derived data.
- 3- The validation of the noise-logger-based models on the field is minimal. Therefore, further verification is required to enhance the developed models.
- 4- Lack of a scientific definition for leak-size categorization and using an assumption instead.
- 5- Limited range of experiments using vibration signals, as only ductile iron pipelines and PVC pipelines as well as small size pipelines no greater than two inches in diameter are used in the experiments.
- 6- The extent of the Lazy Serpent Algorithm is yet to be tested on a real-life project that contains a vast number of events.
- 7- The current representation of the lazy serpent builds on the assumption that any repair event can be represented using three main criteria. Therefore, events that can be represented with more than three dimensions cannot be used in this version.

- 8- The proposed factors for assessment in the lazy serpent and the genetic algorithm prioritization example - criticality, consequence of failure, and condition – are suggested as placeholders. Those factors require the development of actual models that meet the needs of municipalities. Their use in this thesis was in the form of tools to help provide a proof of concept.
- 9- The comparative study utilized linearly deteriorating models to allow for quick computational analysis by the developed models. This is not the case in the real world. Therefore, the exploration of models that deteriorate in a unique manner is required.
- 10- The genetic algorithm approach requires a substantial amount of time and may not arrive at an optimal solution.
- 11- The extent of aggregation can be studied further for all 3D interaction where a full rendering can be developed for a full envelope.

VII.4 Future Work

In this section, possible areas for improving the currently proposed models are previewed. Besides, approaches to extend the use of the proposed models and widen their application are presented. Here, two main components are discussed: (1) enhancement areas and (2) extension areas.

VII.4.1 Enhancement Areas

Current models can be enhanced to produce better results and have an edge in their targeted fields. The suggested enhancement areas are divided into three main groups: (1) Accelerometer Leak Detection System, (2) Acoustic Leak Detection System and (3) Prioritization Approaches.

1- Accelerometer Leak Detection System

- i- Utilize filtration techniques to improve signal clarity. Due to the sensitivity of accelerometers, a wide variety of external factors can find their way into the signal and disrupt its clarity. Additionally, the signal itself is affected by any gravitational shift leading to a slightly unstable signal. The utilization of white noise filtering, Fourier Transform and other forms of filtering may help by stabilizing the signal and thus providing more consistent results in analysis.
- ii- Utilize a more robust and scientifically-based leak size categorization scheme. In this thesis, the categorization of leak sizes is based on a set of assumptions. Categorization can be improved by a professional survey that would indicate at which percentage of the flow rate a leak is considered average or big. Other approaches can also be used for assessment, such leak losses and leak flow rate. A fuzzy approach can also be useful in improving the categorization scheme.

2- Acoustic Leak Detection System

- i- Utilize acoustic filtration techniques and study the effect of each filtration approach: In this study, the acoustic signals are raw and unaltered via filtration, to study the potential of the system in its current state. Meanwhile, eliminating specific frequencies that have little to no effect on the system can be beneficial to get the pure sound produced by the leak only. Those techniques include white noise filtering, impulse response filtering, discrete time filtering and linear filtering.
- ii- Validate the proposed pump identification model further through experiments: In this research, acoustic leak detection models are validated by using the leaks in

the area. The leaks are analyzed via software and assessed on the field. The results of both assessments are compared. Due to the extensive time required to conduct field verification, very few assessments are done. Therefore, further validation of leaks over the span of one year can verify the obtained results and assist in improving the currently developed models.

- iii- Utilize the current leak pinpointing model in other field tests to create a more valid assessment: Similar to the pump identification model, the leak pinpointing model requires further application over the span of one year to collect additional data and improve the model with new data.

3- *Prioritization Approaches*

- i- Utilize a more advanced search technique in the Lazy Serpent Algorithm: The lazy serpent is at an elementary stage currently. It can be improved in terms of speed and capabilities by adding dynamic programming and other advanced problem solving and modeling techniques.
- ii- Explore different genetic algorithm formulations that can speed up the model's convergence: One of the problems of using the genetic algorithm for prioritization is the extensive amount of time required for solving a problem. It is possible to improve the speed of the algorithm by changing the mathematical presentation of the problem or the boundary search approach.

VII.4.2 Extension Areas

Extension areas include extra work to push the developed models and improve their outcome. Similarly, the extension areas are divided into three main categories, as (1) Accelerometer Leak Detection System, (2) Acoustic Leak Detection System and (3) Prioritization Approaches.

1- Accelerometer Leak Detection System

- i- Perform experiments on new materials other than ductile iron and PVC.
- ii- Perform experiments on pipelines greater than 1 inch and 2 inches in size.
- iii- Apply the pinpointing model in large-scale experiments.
- iv- Study the impact of the proposed sensor distribution approach in a network setup.

2- Acoustic Leak Detection System

- i- Study the impact of non-straight connection between pipelines on leak detection accuracy and the leak signal.
- ii- Combine the proposed models with triangulation algorithms.
- iii- Study the possibility of pinpointing using a single sensor or more than two sensors.

3- Prioritization Approaches

- i- Apply the Lazy Serpent Algorithm on a real existing project and compare it against the genetic algorithm.

REFERENCES

- Al Hawari, A., Khader, M., Zayed, T., and Moselhi, O. (2015). "Non-Destructive Visual-Statistical Approach to Detect Leaks in Water Mains." *World Academy of Science, Engineering and Technology, International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering*, 9(3), 230–234.
- Al Hawari, A., Khader, M., Zayed, T., and Moselhi, O. (2016). "Detection of Leaks in Water Mains Using Ground Penetrating Radar." *World Academy of Science, Engineering and Technology, International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering*, 10(4), 422–425.
- Alkassseh, J., Adlan, M. N., Abustan, I., Aziz, H. A., and Hanif, A. B. M. (2013). "Applying minimum night flow to estimate water loss using statistical modeling: A case study in Kinta Valley, Malaysia." *Water resources management*, Springer, 27(5), 1439–1455.
- Almeida, F., Joseph, M. J. B. P. F., Dray, S., Witfield, S., and Paschoalini, A. T. (2014). "An investigation into the effects of resonances on the time delay estimate for leak detection in buried plastic water distribution pipes." *Eurodyn 2014: IX International Conference On Structural Dynamics*, 3129–3136.
- Atef, A., Zayed, T., Hawari, A., Khader, M., and Moselhi, O. (2016). "Multi-tier method using infrared photography and GPR to detect and locate water leaks." *Automation in Construction*, Elsevier, 61, 162–170.
- Awawdeh, A., Bukkapatnam, S., Kumara, S., Bunting, C., and Komanduri, R. (2006). "Wireless sensing of flow-induced vibrations for pipeline integrity monitoring." *2006 IEEE Sensor Array and Multichannel Signal Processing Workshop Proceedings, SAM 2006*, 114–117.
- AWWA. (1987). *Leaks in water distribution systems: a technical/economic overview*. AWWA,

Denver, Colorado, USA.

- Belouchrani, A., Amin, M. G., Thirion-Moreau, N., and Zhang, Y. D. (2013). "Source separation and localization using time-frequency distributions: an overview." *IEEE Signal Processing Magazine*, IEEE, 30(6), 97–107.
- Belsito, S., Lombardi, P., Andreussi, P., and Banerjee, S. (1998). "Leak detection in liquefied gas pipelines by artificial neural networks." *AIChE Journal*, 44(12), 2675–2688.
- Billmann, L., and Isermann, R. (1987). "Leak detection methods for pipelines." *Automatica*, Elsevier, 23(3), 381–385.
- Brennan, M., Gao, Y., and Joseph, P. (2007). "On the relationship between time and frequency domain methods in time delay estimation for leak detection in water distribution pipes." *Journal of Sound and Vibration*, Elsevier, 304(1), 213–223.
- Cai, X., and Li, K. N. (2000). "Genetic algorithm for scheduling staff of mixed skills under multi-criteria." *European Journal of Operational Research*, 125(2), 359–369.
- Cataldo, A., Persico, R., Leucci, G., De Benedetto, E., Cannazza, G., Matera, L., and De Giorgi, L. (2014). "Time domain reflectometry, ground penetrating radar and electrical resistivity tomography: a comparative analysis of alternative approaches for leak detection in underground pipes." *NDT & E International*, Elsevier, 62, 14–28.
- Cheong, L. C. (1991). "Unaccounted for water and economics of leak detection. 8th Int." *Proc. International Water Supply Congress and Exhibition*, International Water Supply Association, Copenhagen, 11–16.
- Colombo, A. F., and Karney, B. W. (2002). "Energy and Costs of Leaky Pipes: Toward Comprehensive Picture." *Journal of Water Resources Planning and Management*, 128(6), 441–450.

- Colorni, A., Dorigo, M., and Maniezzo, V. (1992b). “A genetic algorithm to solve the timetable problem.” *Politecnico di Milano, Milan, Italy TR*, 60–90.
- Cortes, C., and Vapnik, V. (1995). “Support-vector networks.” *Machine learning*, Springer, 20(3), 273–297.
- Covas, D., Ramos, H., and De Almeida, A. B. (2005). “Standing wave difference method for leak detection in pipeline systems.” *Journal of Hydraulic Engineering*, American Society of Civil Engineers, 131(12), 1106–1116.
- Datamatic Inc. (2008). “Permalog+ Leak Noise Loggers, used in conjunction with a MOSAIC Mesh Network or ROADRUNNER Mobile & Handheld data collection platforms, are today’s most powerful tool to protect precious resources and guard against costly and damaging water leaks.” *Datamatic Inc.*, <www.datamatic.com> (Mar. 3, 2015).
- Datta, S., and Sarkar, S. (2016). “A review on different pipeline fault detection methods.” *Journal of Loss Prevention in the Process Industries*, Elsevier, 41, 97–106.
- Davis, S. (2003). “Priority Algorithms.” *Priority Algorithms*, <http://cseweb.ucsd.edu/~sdavis/res_exam.pdf> (Jan. 24, 2017).
- Echologies Inc. (2006). “Operation Management Training.” *Department of Environment and Conservation Newfoundland Canada*, <http://www.env.gov.nl.ca/env/waterres/training/adww/manageoperation/11_m_bracken_gander_moe.pdf> (Nov. 27, 2016).
- El-Abbasy, M. S., Mosleh, F., Senouci, A., Zayed, T., and Al-Derham, H. (2016). “Locating Leaks in Water Mains Using Noise Loggers.” *Journal of Infrastructure Systems*, American Society of Civil Engineers, 4016012.
- El-Abbasy, M. S., Senouci, A., Zayed, T., Asce, M., and Mirahadi, F. (2014). “Condition Prediction Models for Oil and Gas Pipelines Using Regression Analysis.” 140(6), 1–17.

- El-Zahab, S., Asaad, A., Abdelkader, E. M., and Zayed, T. (2017). "Collective thinking approach for improving leak detection systems." *Smart Water*, 2(1), 3.
- El-Zahab, S., Mosleh, F., Zayed, T., Zahab, S. El, Mosleh, F., and Zayed, T. (2016). "An Accelerometer-Based Real-Time Monitoring and Leak Detection System for Pressurized Water Pipelines." *Pipelines 2016*, 257–268.
- Elbehairy, H., Elbeltagi, E., Hegazy, T., and Soudki, K. (2006). "Comparison of two evolutionary algorithms for optimization of bridge deck repairs." *Computer-Aided Civil and Infrastructure Engineering*, 21(8), 561–572.
- Eyuboglu, S., Mahdi, H., Al-Shukri, H., and Rock, L. (2003). "Detection of water leaks using ground penetrating radar." *3rd International Conference on Applied Geophysics 2003*.
- Fahmy, M., and Moselhi, O. (2009). "Automated detection and location of leaks in water mains using infrared photography." *Journal of Performance of Constructed Facilities*, American Society of Civil Engineers, 24(3), 242–248.
- Fantozzi, M., Calza, F., and Lambert, A. (2009). "Experience and Results Achieved in Introducing District Metered Areas (DMA) and Pressure Management Areas (PMA) at Enia Utility (Italy)." *Proceedings of the 5th IWA Water Loss Reduction Specialist Conference*, 153–160.
- Fletcher, T. (2009). "Support vector machines explained." *Tutorial Paper*, <[http://www.tristanfletcher.co.uk/SVM Explained.pdf](http://www.tristanfletcher.co.uk/SVM%20Explained.pdf)> (Jan. 25, 2017).
- Gao, Y., Brennan, M. J., Joseph, P. F., Muggleton, J. M., and Hunaidi, O. (2005). "On the selection of acoustic/vibration sensors for leak detection in plastic water pipes." *Journal of Sound and Vibration*, 283(3–5), 927–941.
- Geiger, G., Vogt, D., and Tetzner, R. (2006). "State-of-the-Art in Leak Detection and

- Localisation 1 Regulatory Framework.” *Oil Gas European Magazine*, URBAN VERLAG, 32(4), 193.
- Ghazali, M., Staszewski, W., Shucksmith, J., Boxall, J., and Beck, S. (2011). “Instantaneous phase and frequency for the detection of leaks and features in a pipeline system.” *Structural Health Monitoring*, 10(4), 351–360.
- Giustolisi, O., and Berardi, L. (2009). “Prioritizing Pipe Replacement: From Multiobjective Genetic Algorithms to Operational Decision Support.” *Journal of Water Resources Planning and Management*, 135(6), 484–492.
- Goldberg, D. E. (2006). *Genetic Algorithms in Search, Optimization, and Machine LEarning*. Pearson Education, Upper Saddle River, New Jersey, USA.
- Gurney, K. (1997). *An introduction to neural networks*. CRC press, Boca Raton, Florida, USA.
- Hamilton, S. (2009). “ALC in Low Pressure Areas - It can be Done.” *Proceedings of Water Loss 2009 South Africa*, CapeTown, South Africa, 131–137.
- Hauge, E., Aamo, O. M., and Godhavn, J.-M. (2007). *Model Based Pipeline Monitoring With Leak Detection. IFAC Proceedings Volumes*, IFAC.
- Hossin, M., and Sulaiman, M. N. (2015). “A review on evaluation metrics for data classification evaluations.” *International Journal of Data Mining & Knowledge Management Process, Academy & Industry Research Collaboration Center (AIRCC)*, 5(2), 1.
- Hunaidi, O., and Wang, A. (2006). “A new system for locating leaks in urban water distribution pipes.” *Management of Environmental Quality: An International Journal*, Emerald Group Publishing Limited, 17(4), 450–466.
- Hunaidi, O., Wang, A., Bracken, M., Gambino, T., and Fricke, C. (2004). “Acoustic methods for locating leaks in municipal water pipe networks.” *International Water Demand*

Management Conference, 1–14.

Inaudi, D., Glisic, B., Figini, A., Walder, R., Belli, R., and Walder, R. (2008). “Pipeline Leakage Detection and Localization Using Distributed Fibre Optic Sensing.” *2008 7th International Pipeline Conference*, 1–8.

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

Kim, D., Ha, J., and You, K. (2011). “Adaptive extended Kalman filter based geolocation using TDOA/FDOA.” *International Journal of Control and Automation*, 4(2), 49–58.

Krchnak, K. (2016). “Water Scarcity.” *Water Scarcity - World Wild Life*, <<https://thewaterproject.org/water-scarcity/>> (Dec. 18, 2016).

KVS. (2015). “Tracer Gas Detection.” *Tracer Gas Detection*, <<http://www.leakdetection-technology.com/science/leak-detection-with-tracer-gas-methods>> (Mar. 9, 2016).

Lay-Ekuakille, A., Vendramin, G., and Trotta, A. (2009). “Spectral analysis of leak detection in a zigzag pipeline: A filter diagonalization method-based algorithm application.” *Measurement*, Elsevier, 42(3), 358–367.

Lay-Ekuakille, A., Vergallo, P., and Trotta, A. (2010). “Impedance method for leak detection in zigzag pipelines.” *Measurement Science Review*, 10(6), 209–213.

Lee, P. J., V'itkovsk'y, J. P., Lambert, M. F., Simpson, A. R., and Liggett, J. A. (2005). “Frequency domain analysis for detecting pipeline leaks.” *Journal of Hydraulic Engineering*, American Society of Civil Engineers, 131(7), 596–604.

Li, W., Ling, W., Liu, S., Zhao, J., Liu, R., Chen, Q., Qiang, Z., and Qu, J. (2011). “Development of systems for detection, early warning, and control of pipeline leakage in

- drinking water distribution: A case study.” *Journal of Environmental Sciences*, The Research Centre for Eco-Environmental Sciences, Chinese Academy of Sciences, 23(11), 1816–1822.
- Liggett, J. A., and Chen, L.-C. (1994). “Inverse transient analysis in pipe networks.” *Journal of Hydraulic Engineering*, American Society of Civil Engineers, 120(8), 934–955.
- Lobato de Almeida, F. C. (2013). “Improved Acoustic Methods for Leak Detection in Buried Plastic Water Distribution Pipes.” University of South Hampton.
- Mandal, S. K., Chan, F. T. S., and Tiwari, M. K. (2012). “Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM.” *Expert Systems with Applications*, Elsevier Ltd, 39(3), 3071–3080.
- Martini, A., Troncossi, M., and Rivola, A. (2015). “Automatic leak detection in buried plastic pipes of water supply networks by means of vibration measurements.” *Shock and Vibration*, 2015, 1–13.
- Mashford, J., De Silva, D., Marney, D., and Burn, S. (2009). “An approach to leak detection in pipe networks using analysis of monitored pressure values by support vector machine.” *Third International Conference on Network and System Security, 2009. NSS’09.*, 534–539.
- MEMS and Nanotechnology Exchange. (2015). “What is MEMS Technology?” <<https://www.mems-exchange.org/MEMS/what-is.html>> (Feb. 26, 2016).
- Meng, L., Yuxing, L., Wuchang, W., and Juntao, F. (2012). “Experimental study on leak detection and location for gas pipeline based on acoustic method.” *Journal of Loss Prevention in the Process Industries*, Elsevier Ltd, 25(1), 90–102.
- Meniconi, S., Brunone, B., and Ferrante, M. (2010). “In-line pipe device checking by short-period analysis of transient tests.” *Journal of Hydraulic Engineering*, American Society of

- Civil Engineers, 137(7), 713–722.
- Mitchell, M. (1998). “L.D. Davis, handbook of genetic algorithms.” *Artificial Intelligence*, 100(1–2), 325–330.
- Morcous, G., and Lounis, Z. (2005). “Maintenance optimization of infrastructure networks using genetic algorithms.” *Automation in Construction*, 14(1), 129–142.
- Moselhi, O., and Hassanein, a. (2003). “Optimized Scheduling of Linear Projects.” *Journal of Construction Engineering and Management*, 129(6), 664–673.
- Mostafapour, a., and Davoudi, S. (2013). “Analysis of leakage in high pressure pipe using acoustic emission method.” *Applied Acoustics*, Elsevier Ltd, 74(3), 335–342.
- Mounce, S. R., and Machell, J. (2006). “Burst detection using hydraulic data from water distribution systems with artificial neural networks.” *Urban Water Journal*, 3(1), 21–31.
- Mpesha, W., Gassman, S. L., and Chaudhry, M. H. (2001). “Leak detection in pipes by frequency response method.” *Journal of Hydraulic Engineering*, American Society of Civil Engineers, 127(2), 134–147.
- Müller, B., Reinhardt, J., and Strickland, M. T. (2012). *Neural networks: an introduction*. Springer Science & Business Media, Berlin, Germany.
- Piller, O., and Van Zyl, J. E. (2014). “Incorporating the FAVAD leakage equation into water distribution system analysis.” *Procedia Engineering*, 613–617.
- Pudar, R. S., and Liggett, J. A. (1992). “Leaks in Pipe Networks.” *Journal of Hydraulic Engineering*, 118(7), 1031–1046.
- Puretech Ltd. (2015). “Smartball.” <http://www.puretechltd.com/products/smartball/smartball_leak_detection.shtml> (Feb. 26, 2016).
- Rajani, B., and Kleiner, Y. (2001). “Comprehensive review of structural deterioration of water

- mains: Physically based models.” *Urban Water*, 3(3), 151–164.
- Rawlings, J. O., Pantula, S. G., and Dickey, D. a. (1998). *Applied Regression Analysis: A Research Tool. Technometrics*.
- Romano, M., Woodward, K., and Kapelan, Z. (2017). “Statistical Process Control Based System for Approximate Location of Pipe Bursts and Leaks in Water Distribution Systems.” *Procedia Engineering*, Elsevier, 186, 236–243.
- Royal, A. C. D., Atkins, P. R., Brennan, M. J., Chapman, D. N., Chen, H., Cohn, A. G., Foo, K. Y., Goddard, K. F., Hayes, R., Hao, T., Lewin, P. L., Metje, N., Muggleton, J. M., Naji, A., Orlando, G., Pennock, S. R., Redfern, M. A., Saul, A. J., Swinger, S. G., Wang, P., and Rogers, C. D. F. (2011). “Site Assessment of Multiple-Sensor Approaches for Buried Utility Detection.” *International Journal of Geophysics*, Hindawi Publishing Corporation, 2011, 1–19.
- Sala, D., and Kołakowski, P. (2014). “Detection of leaks in a small-scale water distribution network based on pressure data - Experimental verification.” *Procedia Engineering*, 1460–1469.
- Santos, R. B., Rupp, M., Bonzi, S. J., and Fileti, A. M. F. (2013). “Comparison between multilayer feedforward neural networks and a radial basis function network to detect and locate leaks in pipelines transporting gas.” *Chemical Engineering Transactions*, 32, 1375–1380.
- Schempf, H., Mutschler, E., Goltsberg, V., Skoptsov, G., Gavaert, A., and Vradis, G. (2003). “Explorer: Untethered real-time gas main assessment robot system.” *Proc. of Int. Workshop on Advances in Service Robotics, ASER*.
- Shibley, J. A. (2013). “Enhanced Sonar Array Target Localization Using Time-Frequency

Interference Phenomena.”

- Shinozuka, M., Chou, P. H., Kim, S., Kim, H. R., Yoon, E., Mustafa, H., Karmakar, D., and Pul, S. (2010). “Nondestructive monitoring of a pipe network using a MEMS-based wireless network.” *Proceedings of SPIE - Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security*, 76490P.
- Silva, R. a., Buiatti, C. M., Cruz, S. L., and Pereira, J. a. F. R. (1996). “Pressure wave behaviour and leak detection in pipelines.” *Computers & Chemical Engineering*, 20(96), S491–S496.
- Sokolova, M., and Lapalme, G. (2009). “A systematic analysis of performance measures for classification tasks.” *Information Processing & Management*, Elsevier, 45(4), 427–437.
- Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H. B., and Whittle, A. J. (2013). “Wavelet-based burst event detection and localization in water distribution systems.” *Journal of Signal Processing Systems*, 72(1), 1–16.
- Stampolidis, A., Soupios, P., Vallianatos, F., and Tsokas, G. N. (2003). “Detection of leaks in buried plastic water distribution pipes in urban places - A case study.” *Proceedings of the 2nd International Workshop on Advanced Ground Penetrating Radar*, 120–124.
- Stoianov, I., Nachman, L., Madden, S., Tokmouline, T., and Csail, M. (2007b). “PIPENET: A wireless sensor network for pipeline monitoring.” *Information Processing in Sensor Networks, 2007. IPSN 2007. 6th International Symposium on*, 264–273.
- Sun, Z., Wang, P., Vuran, M. C., Al-Rodhaan, M. A., Al-Dhelaan, A. M., and Akyildiz, I. F. (2011). “MISE-PIPE: Magnetic induction-based wireless sensor networks for underground pipeline monitoring.” *Ad Hoc Networks*, Elsevier B.V., 9(3), 218–227.
- Sykes, A. O. (2007). “An Introduction to Regression Analysis.” *American Statistician*, 61(1), 101–101.

- United States Environmental Protection Agency. (2009). *Control And Mitigation of Drinking Water Losses in Distribution Systems*. Washington, DC.
- Van Eck, N. J., and Waltman, L. (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping." *Scientometrics*, Springer, 84(2), 523–538.
- Van Eck, N. J., and Waltman, L. (2014). "CitNetExplorer: A new software tool for analyzing and visualizing citation networks." *Journal of Informetrics*, Elsevier, 8(4), 802–823.
- Varone, S. ;, and Varsalona, P. (2012). "Detecting Leaks with Infrared Thermography." *Habitat Magazine*, New York, NY, 48–50.
- Verde, C., Molina, L., and Torres, L. (2014). "Parameterized transient model of a pipeline for multiple leaks location." *Journal of Loss Prevention in the Process Industries*, Elsevier Ltd, 29(1), 177–185.
- Wang, X.-J., Lambert, M. F., Simpson, A. R., Liggett, J. A., V \i' tkovsk\y, J. P., Asce, M., Liggett, J. A., and V1, J. P. (2002). "Leak detection in pipelines using the damping of fluid transients." *Journal of Hydraulic Engineering*, American Society of Civil Engineers, 128(7), 697–711.
- WCT Products. (2015). "Vivax Vcam5." <<http://www.wctproducts.com/products/pipe-cameras/vivax-metrotech-vcam-5-camera.php>> (Mar. 5, 2016).
- Whittle, A. J., Girod, L., Preis, A., Allen, M., Lim, H. B., Iqbal, M., Srirangarajan, S., Fu, C., Wong, K. J., and Goldsmith, D. (2010). "WaterWiSe@SG: A Testbed for Continuous Monitoring of the Water Distribution System in Singapore." *Water Distribution Systems Analysis 2010*, 1362–1378.
- Zielke, W. (1968). "Frequency-dependent friction in transient pipe flow." *Journal of basic engineering*, American Society of Mechanical Engineers, 90(1), 109–115.

Van Zyl, J. E., and Clayton, C. R. I. (2007). "The effect of pressure on leakage in water distribution systems." *Proceedings of the Institution of Civil Engineers-Water Management*, 109–114.

APPENDIX A: LEAK DETECTION TECHNOLOGIES

Multiple technologies have been developed throughout the years to help identify and locate leaks within water networks. In this review, the modern state-of-the-art techniques will be reviewed, and their advantages and disadvantages will be pointed out.

A.1 Listening Devices

Both electrical and mechanical geophones are used to listen to buried water pipelines from the surface. These devices are accurate and highly sensitive that they can detect the exact location of the leak, and also cheap to purchase and easy to set up.

The accuracy of geophones depends highly on the proficiency and the experience of the operator, and it also might fail to detect some leak classes. Furthermore, the exact location of the pipeline to be assessed must be marked so that the operator would know where to put the device. The examination renders the area above the pipe unusable in case it is a street or highly utilized area.

Similar to Geophones, Hydrophones try to listen to leaks by sometimes being placed in the system and rarely on the surface of the ground. Hydrophones can be more accurate than geophones in detecting leaks but they require more training than geophones to operate, and they are approximately seven times more expensive than geophones (United States Environmental Protection Agency 2009).

To detect a leak, these devices rely on the high-frequency acoustic signals sent by the release of pressurized fluids, to detect leak existence and leak locations. Sound frequencies are then amplified and filtered at 1 kHz using a preamplifier to remove high-frequency noises that are not related to the network. By measuring the time delay between two detection instants between two given listeners the leak can be pinpointed by relating propagation speed within the medium with time and distance.

A.2 Leak Noise Loggers

Leak Noise loggers are placed in utility holes without any trenching or drilling; they can be used as a permanent monitoring, semi-permanent monitoring, or leak surveying technique. Noise Loggers operate by implementing sophisticated algorithms to identify the sound emitted by normal operations compared to that of the leak, thus identifying leaks immediately as they occur. Also, this technology is automatic thus eliminating human error. Noise loggers also have low maintenance and battery replacement cost for long-term use.

This technology has a very high initial cost for a real-time monitoring system, and it does not identify the exact leak location without the use of correlators (Datamatic Inc. 2008).

A.3 Infrared Thermography

According to Fahmy and Moselhi (2009) “Thermography (IR) camera measures and images the emitted infrared radiation from an object. It can detect thermal contrasts on pavement surface due to water leaks.”

This technology has been tested on non-metallic pipelines and has shown accurate results, yet these results were not as accurate as the results provided by acoustic technologies such as geophones and are unreliable under the cases of damaged pipelines. Furthermore, IR Camera requires marking of the pipeline on the ground surface so that the machine can move above the pipeline. Also the machine is profoundly affected by the surrounding weather and any variations in soil conditions and temperature.

The infrared technology relies on the energy released by the vibration and motion of particles that release energy emissions based on their temperature. Infrared energy is not visible to the naked eye. Infrared thermography utilizes wavelengths that are limited to the electromagnetic

range between 0.4 and 0.7 micrometers. Infrared technologies detect wavelength ranges larger than 0.7 micrometers, thus detecting the temperature distribution inside the pipe and in the surrounding environment to identify any temperature anomalies that might indicate the existence of a leak as shown in Figure A-1 (Varone and Varsalona 2012).

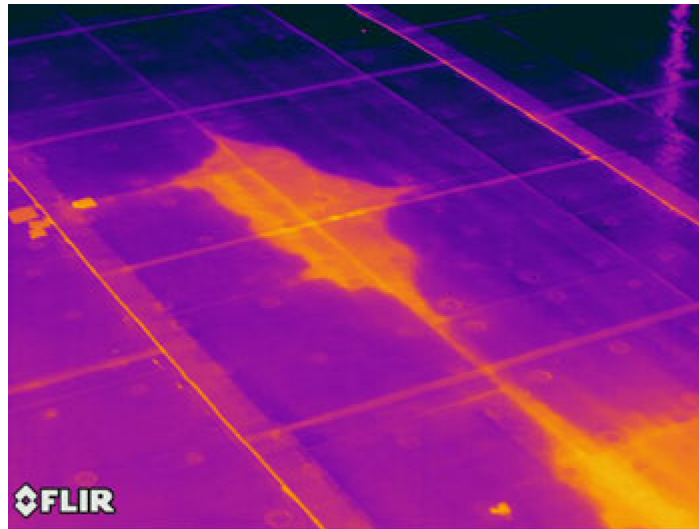


Figure A-1: Infrared image of a floor showing the distinctive color of entrapped moisture

A.4 Tracer Gas

Tracer gasses is a leak detection technique that utilizes pressurizing nontoxic and insoluble gasses into leaks, these gasses contain ammonia, halogens, and helium, where helium is the most sensitive.

Given that the utilized gasses are lighter than air they will tend to go out through leaks and then seep out through the soil or pavements. Later on, these gasses are traced and detected using a man operated detector to identify the locations of leaks through detecting the seepage of tracer gasses (KVS 2015).

A.5 Ground Penetrating Radar

The Ground Penetrating Radar (GPR) technology utilizes electromagnetic waves (between 125 Mhz and 370 MHz) and transmits them into the ground to identify leak location via imaging the sub-terrain including the pipe and the leak.

The advantages of this technology lie in its capability to detect leaks regardless of the material of the pipe, for any diameter size above one inch and reaching to a depth of 4 meters. Furthermore, this device can be easily transported between sites, and it does not require a lot of experience or training on behalf of the operator to operate it.

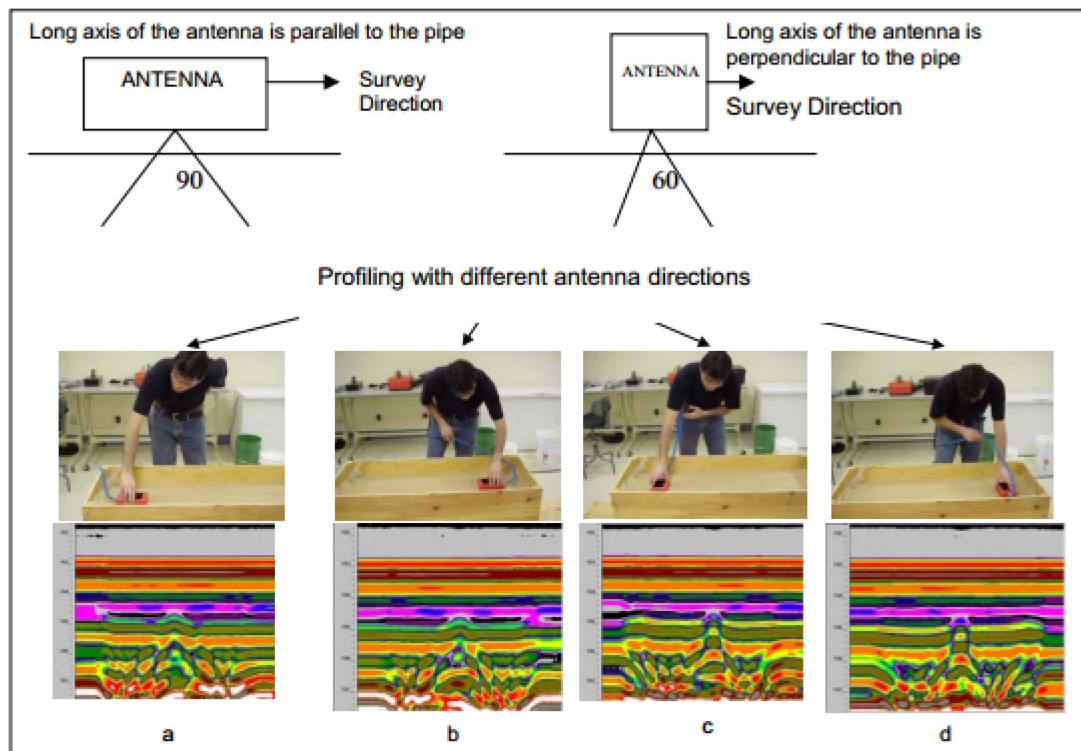


Figure A-2: GPR Leak detection of a simulated leak at multiple angles. (Eyuboglu et al. 2003)

On the other hand, this technology has multiple disadvantages namely requiring access to the road above the pipeline – thus disturbing traffic –, experience and training are required to accurately indicate the position of the leak and the dependency on the pipeline’s bedding and

surrounding conditions. Also, this technology is expensive where the machine price can range between 15,000\$ to 31,000\$ (United States Environmental Protection Agency 2009).

Eyuboglu et al. (2003) developed a mathematical model that utilizes the amplitude of radar reflection to visualize the state of the pipe and detect precisely where the leak occurred. Figure A-2 illustrates how the GPR is moved above the soil to detect the condition of the pipeline.

A.6 Leak Detecting Robots

Multiple robotic devices were developed to perform in pipe inspection and determine leak locations in sewers. These devices can be wireless devices or cord connected. Furthermore, some leak detection robots can also perform leak repair tasks.

A. Smart-Ball

Smart-Ball is a free-swimming technology developed to detect leaks from within live large water pipelines. Smart-ball technology is composed of a foam ball with an aluminum alloy core, within the aluminum core, a highly sensitive detection instrument is placed.

Smart-ball does not create any noise when passing through the pipeline. Therefore, it can detect tiny leaks. Also, the Smart-ball has a location accuracy within 3 meters of estimated leak location, it is very flexible due to its small size and can enter multiple sizes of pipelines. The Smart-ball requires two points of access to assess a pipe, a point for insertion and a point for extraction, and it is a non-destructive technology for leak detection.

On the other hand, Smart-ball can only be operated by the manufacturing company only, and the ball might divert from the path it was required to search, or even the ball might get stuck (Puretech Ltd. 2015).

PureTech Limited utilizes the acoustic sensor within the smart-ball device to listen to all the sounds emitted inside the vicinity of the pipe. Furthermore, PureTech utilizes their software to

analyze the sounds they are hearing and identify the locations of leaks, valves, as well as air pockets. Figure A-3 Further illustrates a smart ball passing through a pipeline and its possible outputs.

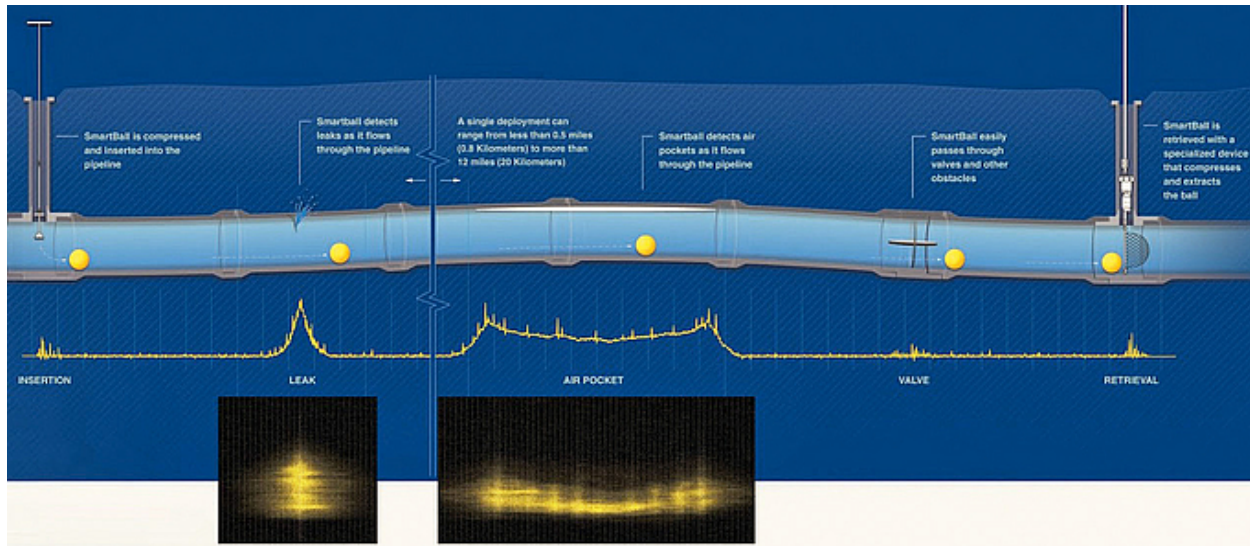


Figure A-3: Movement of Smart-Ball within a pipeline with possible results.

(Puretech Ltd. 2015)

B. Explorer

Schempf et al. developed a new photogrammetric leak detection robot named Explorer. Explorer is a robot developed mainly to detect leaks in gas pipelines; it is made of seven compartments that are composed of 2 locomotors with camera heads, two batteries, two support compartments, and one computer compartment in the middle as shown in Figure A-4.

Explorer travels inside 6-inch and 8-inch pipelines as they operate with scanning the entire network with its long battery life and taking a picture of the sections they pass through using their front and back camera. The images are taken via fish-eye camera lenses and then processed by the computer within Explorer. These images are transferred to the operator's computers where Laplace enhancements are performed on the images and then delivered to the operator for visual inspection (Schempf et al. 2003).

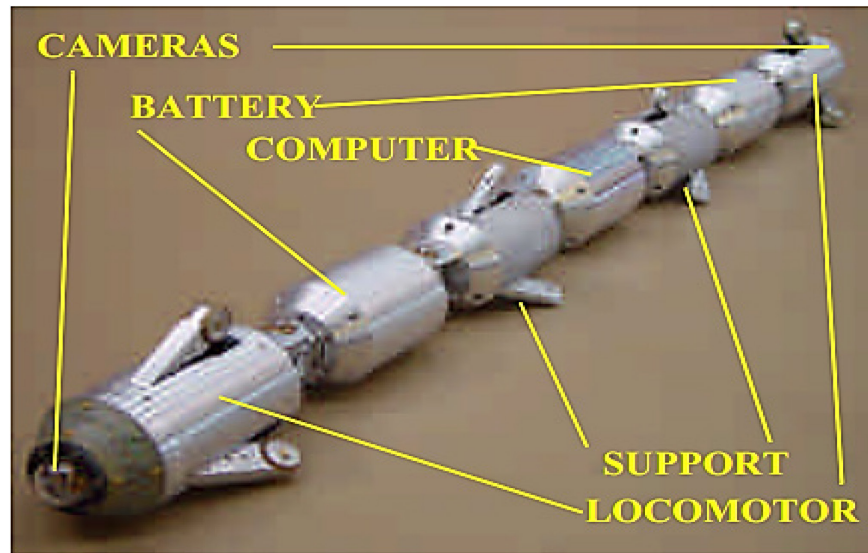


Figure A-4: Overall Shape and architecture of Explorer leak detection robot.

(Schempf et al. 2003)

C. Beaver

DSI Robotics developed a cord attached robot that can be inserted into pipelines and sewers for pipeline assessment, leak detection, and leak repair purposes. The system is known as Beaver; the robot is attached to a machine that helps control the inspection robot and view what it is seeing.

Beaver relies on Closed Circuit Television (CCTV) that is built into the robot. The robot compartment is inserted into the pipeline that requires studying and takes images that are processed in the stationary device to be viewed by the operator. Figure A-5 shows the robot device and the van that is attached to it.



Figure A-5: Beaver robot along with the van that operates and displays the robot. Photo courtesy of DSI Robotics

D. Pear-point

Much like Beaver, pear point is a tethered device used to assess live pipelines. Pear-point utilizes a camera with a long tube to transfer live images to the monitor attached to the reeling device – as shown in Figure A-6 – for the operator to identify the locations of leaks within the pipeline (WCT Products 2015).



Figure A-6: Pear-point leak detection system. Courtesy of WCT Products.

A.7 Wireless Micro-Electro-Mechanical Systems (MEMS)

Micro-Electro-Mechanical Systems are microfabricated mechanical and electro-mechanical devices and structures. MEMS are usually composed of four main elements:

- Micro-Sensors
- Micro-Actuators
- Micro-Electronics
- Micro-Structures (MEMS and Nanotechnology Exchange 2015).

Multiple types of MEMS were used in leak detection of water mains mainly accelerometers, acoustic, and thermal.

MEMS technology provides continuous water network monitoring for any leaks from the moment they are placed. MEMS's accelerometers and data gathering hub are relatively cheap compared to other technologies and very accurate.

On the other hand, MEMS application is still mostly theoretical and requires further research. Furthermore, MEMS needs further testing on long pipes and some pipeline materials as well (Kim et al. 2011).