Intelligent One-Point Damage Localization of An Isotropic Surface Pipeline Using Gaussian Process Regression

Shida Khazaeli

A thesis
in
The Department
of
Building, Civil, and Environmental Engineering

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Applied Science
Concordia University
Montréal, Québec, Canada

September 2018

Concordia University
School of Graduate Studies

This is to certify that the thesis prepared

By: **Shida Khazaeli**

Entitled: **Intelligent One-Point Damage Localization of An Isotropic Surface Pipeline Using Gaussian Process Regression**

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

Master of Applied Science

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

Signed by the final examining commitee:

Fuzhan Nasiri _____ Chair

Lucia Tirca _____ Examiner

Amin Hammad _____ Examiner

Ashutosh Bagchi _____ Supervisor

Approved _____
Chair of Department or Graduate Program Director

_____ 20 _____ _____

Amir Asif , Dean
Faculty of Engineering and Computer Science

Abstract

Intelligent One-Point Damage Localization of An Isotropic Surface Pipeline Using Gaussian Process Regression

Shida Khazaeli

Pipelines are subjected to many damaging agents, such as, earthquake, ground movement, and aging which are responsible for important financial expenses. Structural Health Monitoring (SHM) of civil structures using arrays of sensors is promising such that data form the monitoring systems enable us to trace the structural anomalies and performance for early treatments. The need for introducing faster and intelligent methods has helped researchers propose novel approaches for such monitoring procedures. In this study a new method is introduced for monitoring of surface pipelines used primarily for oil and gas. The framework takes the advantage of Gaussian Process Regression Method (GPRM) to create a probabilistic predictive model for damage detection and the subsequent localization of the defect. To this end, an isotropic pipeline is modeled numerically and validated with an experimental setup. Afterwards, the model is extended to the real-life application to establish a meta model. Damages are introduced as small holes at different locations (one at each time). The GPRM is used to map the system responses to the selected statistical features which are utilized as indicators for the existence of the damages and their locations. GPRM reveals more promising results compared with conventional regression analysis. It considers the uncertainties due to lack of observation. In addition, it is an updatable approach with having local effects on the model. In another words, it affects the model in the vicinity of new observations. Moreover, among selected statistical features, number of peaks greater than or equal to 20% and 60% of the maximum peak values show better results corresponding to damage localization. Also the curve length and correlation coefficient of the system response (induced signal) are found to be efficient for damage detection. The novel method has been validated with filed measurements and experimental data and found to work efficiently.

Acknowledgments

"Predicting the future isn't magic, it's artificial intelligence."
- Dave Waters

Without a shadow of doubt, this thesis would not have been accomplished without the guidance and help of many individuals who contributed their valuable assistance and advice in carrying out and completion of the current study.
Above all, none of this would have been possible without the support of my family and their patience during the study. My familiy to whom this thesis is dedicated to, undoubtedly has been the source of love, support, and strength throughout all these years. I would love to express my utmost gratitute towards them.
I would love to express my deepest gratitute to **Professor Bagchi**, for his guidance and patience during the study as well as encouraging me carrying out this research. Also, I would love to thank all students who have been helping me in one way or another during the study.

Contents

List of Figures viii

List of Tables xi

1 Introduction 1

1.1 Background and motivation . . . 1
1.2 Research objectives, and contribution . . . 3
1.3 Thesis structure . . . 3

2 Literature review 5

2.1 Methods of health monitoring for pipeline . . . 5
2.2 Causes of gas leakage in pipelines and leak detection methods . . . 6
2.2.1 Causes of gas leakage . . . 6
2.2.2 Leak detection methods . . . 7
2.2.3 Leakage classification . . . 8
2.2.4 Gas detectors . . . 9
2.3 Summary . . . 10

3 Experiment and frequency domain analysis 11

3.1 Pipeline experiment setup . . . 11
3.1.1 Hammer test . . . 12
3.1.2 Data processing and mode shapes . . . 13
3.2 Frequency domain analysis and validation . . . 14
3.2.1 Frequency domain analysis . . . 14
3.2.2 Numerical model validation . . . 15
3.3 Summary . . . 16

4 Intelligent damage detection detection and localization 18

4.1 Intelligent framework . . . 18
4.2 Numerical Simulation . . . 20
4.2.1 Spatial characteristics . . . 20
4.2.2 Mechanical characteristics . . . 20

4.2.3 Loading characteristics . . . 20
4.2.4 Meshing . . . 21
4.2.5 Dynamic simulation . . . 21
4.2.6 Simulation results . . . 22
4.3 Damage library . . . 24
4.4 Features . . . 24
4.4.1 Energy Norm . . . 25
4.4.2 Curve length . . . 25
4.4.3 Local Maximum Feature . . . 25
4.4.4 Kurtosis . . . 26
4.4.5 Correlation Coefficient . . . 26

5 Gaussian Process Regression Method **31**
5.1 Background . . . 31
5.2 Gaussian Process Regression . . . 32
5.2.1 Correlation function . . . 33
5.2.2 Feature selection . . . 35
5.3 Updating GP using noise-free observations . . . 35
5.4 Parameter estimation and model performance . . . 36
5.5 Meta model . . . 37
5.5.1 Gaussian Process for Machine Learning (GPML) tool . . . 37
5.5.2 Over-fitting . . . 38
5.5.3 Illustrative result . . . 38

6 Results, discussion, and comparison **42**
6.1 Results and discussion . . . 42
6.2 Comparison . . . 44
6.3 Decision making . . . 46
6.4 Special study . . . 47
6.5 Summary . . . 50

7 Conclusion and future work **51**
7.1 Conclusion . . . 51
7.2 Future work and recommendation . . . 52

References **54**

Appendices **57**

A Simulated time history data for the pipe **58**
A.1 Section 1 . . . 58

B MATLAB® codes **68**
B.1 Chapter 3 68
B.2 Chapter 4 69
B.3 Chapter 5 70

List of Figures

1 An example of Pig robot for leak detection. Source: *https://sourceable.net* 6
2 An example of ultra sound device for leak detection. Source: *https://www.engineering.com* . . . 7
3 An exampl of surface gas pipeline. Source: *https://www.thomasindcoatings.com* 8
4 FID gas detector device. Source: *http://gasandoil.com.au/* 9
5 Experiment setup 12
6 Schematic experiment setup. 13
7 Normalized mode shapes of free-free pipe. 14
8 First mode shape of the modeld pipeline obtained from ANSYS®. 16
9 Second mode shape of the modeld pipeline obtained from ANSYS®. 17
10 Third mode shape of the modeld pipeline obtained from ANSYS®. 17
11 Fourth mode shape of the modeld pipeline obtained from ANSYS®. 17
12 Intelligent damage localization framework, addopted and modified from [39]. 19
13 Schematic geometry of the modeled pipeline. 20
14 External loading on the modeled pipeline, ANSYS®. 21
15 Internal pressure inside the modeled pipeline, ANSYS®. 21
16 A typicsl mesh grid along the pipe, ANSYS®. 22
17 Multizoning in vicinity of the fault, ANSYS®. 22
18 Induced mesh grid near the fault using multizoning technique, ANSYS®. 23
19 Load monitoring during daynamic analysis (external loading), ANSYS®. 23
20 Load monitoring during daynamic analysis (internal loading), ANSYS®. 24
21 Acceleration response of the pipe in two cases (intact and damaged). 24
22 Abosulte response difference of the pipe in between two cases (intact and damaged). 25
23 $peak \geq 60\%$ 27
24 $peak \geq 20\%$ 27
25 L2. 28
26 CL. 28
27 KU. 29
28 COR. 29
29 MeanPeak. 30
30 Squared Exponential covariance function with different characteristic length-scale. . 34
31 Periodic covariance function with $p = 3$ and different characteristic length-scale. . . . 34

32 Periodic covariance function with $l = 1$ and different peridic length. 35
33 Feature 1. 38
34 GP model for the pipeline with finite observations ID ∈ [1,3,9,15,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0247, l \approx 0]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) = 11.4$ 39
35 GP model for the pipeline with finite observations ID ∈ [1,2,3,4,14,15,16,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0221, l = 1.7577]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) =$ 19.3 40
36 GP model for the pipeline with finite observations ID ∈ [1,2,3,4,5,6,7,8,9,10,11,12,13,15,16,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0194, l = 2.7844]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) =$ 43.7 40
37 GP model for the pipeline with finite observations ID ∈ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0188, l = 2.6545]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) = 46.7$ . . . 41
38 GP model for the pipeline with finite observations ID ∈ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 4.1E3, l = 3.19]^{\mathsf{T}}$. $\ln f(g_i|CL, \sigma_G, l) = 46.7$ 42
39 GP model for the pipeline with finite observations ID ∈ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.42, l = 3.2]^{\mathsf{T}}$. $\ln f(g_i|COR, \sigma_G, l) = 48.45$ 43
40 GP model for the pipeline with finite observations ID ∈ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 13.03, l = 2.13]^{\mathsf{T}}$. $\ln f(g_i|KU, \sigma_G, l) = 45.6$ 43
41 GP model for the pipeline with finite observations ID ∈ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.227, l = 2.78]^{\mathsf{T}}$. $\ln f(g_i|L2, \sigma_G, l) = 43.2$ 44
42 GP model for the pipeline with finite observations ID ∈ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.002, l = 2.8]^{\mathsf{T}}$. $\ln f(g_i|MeanPeak, \sigma_G, l) = 45.8$ 44
43 GP model for the pipeline with finite observations ID ∈ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0063, l = 1.88]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 20\%, \sigma_G, l) = 46.9$ 45
44 Comparison of the regression with three different models for $peak \geq 60\%$. 45
45 Comparison of the regression models and GPRM for $peak \geq 60\%$. 46
46 Larze technology (accelerometer) located on wood 48
47 Larze technology (accelerometer) located on wood 49
48 Fault on the wood beam 49
49 Fault on the wood beam 50
50 GP model for the wood The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 2.6, l = 5, \sigma_V = 0.1]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) = 10.5$ 50
51 Time series of the pipe line with no fault (intact pipe). 58
52 Time series of the pipe line with a fault located located at 3 m. 58
53 Time series of the pipe line with a fault located located at 6 m. 59
54 Time series of the pipe line with a fault located located at 9 m. 59
55 Time series of the pipe line with a fault located located at 12 m. 59
56 Time series of the pipe line with a fault located located at 15.5 m. 59
57 Time series of the pipe line with a fault located located at 18 m. 60
58 Time series of the pipe line with a fault located located at 21 m. 60

59 Time series of the pipe line with a fault located located at 24 m. 60
60 Time series of the pipe line with a fault located located at 27 m. 60
61 Time series of the pipe line with a fault located located at 30.5 m. 61
62 Time series of the pipe line with a fault located located at 33 m. 61
63 Time series of the pipe line with a fault located located at 36 m. 61
64 Time series of the pipe line with a fault located located at 39 m. 61
65 Time series of the pipe line with a fault located located at 42 m. 62
66 Time series of the pipe line with a fault located located at 45.5 m. 62
67 All Frequencies 62
68 Frequency (No Fault vs. Fault at point 3m). 63
69 Frequency (No Fault vs. Fault at point 6m). 63
70 Frequency (No Fault vs. Fault at point 9m). 63
71 Frequency (No Fault vs. Fault at point12m). 64
72 Frequency (No Fault vs. Fault at point 15.5m). 64
73 Frequency (No Fault vs. Fault at point 18m). 64
74 Frequency (No Fault vs. Fault at point 21m). 65
75 Frequency (No Fault vs. Fault at point 24m). 65
76 Frequency (No Fault vs. Fault at point 27m). 65
77 Frequency (No Fault vs. Fault at point 30.5m) 66
78 Frequency (No Fault vs. Fault at point 33m) 66
79 Frequency (No Fault vs. Fault at point 36m). 66
80 Frequency (No Fault vs. Fault at point 39m). 67
81 Frequency (No Fault vs. Fault at point 42m). 67
82 Frequency (No Fault vs. Fault at point 45.5m). 67

List of Tables

1 Mode shape frequency results in proposed experiment 13
2 Mode shape frequency results comparison between experiment and ANSYS®. 16
3 Damage location-covariates table. 26
4 First feature ($peak \geq 60\%$) ID, values, and location of the defect. 39

Chapter 1

Introduction

1.1 Background and motivation

Pipelines play an important role in transporting non-solids such as gas, oil, and etc. Governments and oil and gas firms spend significant amount of money to construct and maintain pipelines for different purposes. There are two primary reasons that pipelines are important. Firstly, pipelines can carry huge amount of liquids for a very long distance with less expenses [34]. For instance, while the capacity of a pipe, which can carry 150,000 barrels/day, it would need 750 trucks per day to carry the same volume [25]. Secondly, with respect to the safety and quick production, pipelines' design addresses the requirement more efficiently [19].

Pipeline netwok are divided into three categories, namely oil, natural gas, and others (water, chemical, and etc) pipelines. While the petroleum pipelines utilize pumps to push oil into the storage facilities, gas transmission pipelines first use compressor to send it for a long distance (up to the *city gate*) and then pressurized the gas for end-users. [34]. Although utilizing pipelines for transporting liquids and gas are beneficial with respect to economic and safety, costs related to them increases become of various incidents, but also by aging that exposes the system for different types of failure. Operation and management of oil and gas lines, the bulk of the cost (more than 50%) in the transfer of oil and gas transportation. Hence, significant financial resources are spent around the world to find new and cheaper ways of monitoring the pipelines to ensure their safety and high performance [26]. Monitoring civil structures using arrays of sensors is promising such that the data enables us to trace the structural anomalies and performance for early treatments. Structural Health Monitoring (SHM) is defined as "a nondestructive in-situ structural evaluation method that uses any of several types of sensors which are attached to, or embedded in, a structure." [10].

There are three major SHM procedures, namely as law-driven (physical-based), data-driven, and hybrid approaches [2]. In physical modeling, the deviation of the structural responses from the normal condition can be measured with the aid of numerical approaches such as Finite Element Method (FEM) [1]. After the initial FE model is built and field measurment, the experimental

data are collected from the sensors and used to update the model (i.e. model calibration). [11]. Afterwards, the models are extended to the large scale models that provide different responses from various induced damages. Finally, by quantifying the damages, engineers can assess the condition of the structures. On the other hand, data-driven methodologies depend on sensed data without any physical modeling. In this context, statistical pattern recognition and machine learning has been of interest of many researcher [8] and [3]. Machine Learning (ML) can be classified as supervised or unsupervised learning approaches. In supervised learning data points are labeled meaning that there is a prior knowledge of the state (intact or damaged) of the structure. There are many approaches for classification of such types of data. The reader is referred to the literature for complete available methods [29] and [15]. However, in many situations, the label of the data are not known (i.e. unsupervised learning or clustering) [29] and [15]. Recently, coupled methods are also reported [27] in order to take the advantages of both approaches. The third approach, hybrid method, is a methodology benefits from two above mentioned approaches [2]. For instance, [18] introduced a lengths computation strategy based on FE models along with Monte-Carlo simulation. The first procedure of SHM procedure is expressed in three levels, namely damage detection, localization, and characterization [10]:

- Level I: The existence of a damage is determined without any further information about location.
- Level II: At this level of SHM presence and location of the damage is identified.
- Level III: Including the previously levels, at this stage type(s) and severity of the damage is determined.

Following [39] the data-driven approach can be identified in six stages.

- Data acquisition: Sensing different types of data under different conditions.
- Preprocessing: Sensed data are preprocessed to improve the signal quality (e.g. filtering and energy-normalizing).
- Feature extraction: Damage-sensitive features from sensed data are extracted at this stage. It is suggested that a large library of features is needed because features may be sensitive to different damages. Such features may be extracted by means of statistics, signal processing as well as machine learning techniques.
- Feature selection: For the sake of computational efficiency and performance of the machine learning algorithms [15] this module examines such features to extract most effective damage-sensitive features.
- Pattern recognition: The main goal of this module is to classify data sets. Classification aims to classify data into different groups based on specific similarities.
- Decision: Having all information from, this module is dedicated to combining all results to allow engineers make a global/local decision(s) for the system.

It is notable that the above-mentioned stages are in fact the development of a four-phase framework previously introduced by [8] and [9].

1.2 Research objectives, and contribution

Primary objectives of this research are emanated from creating a data-driven model for monitoring of the pipelines. The model benefits from lack of need for establishing law-based models in which the characteristics of the complex structure have to be known. Another issue is related to the uncertainty of real-world problem. The application of SHM in real-world contains uncertainties with respect to the phenomena under consideration as well as observations (sensing). As such, the objectives, which results in the contributions, of this study can be stated as follows.

- **Develop new data-driven framework**
  Develop a generic framework for an intelligent data-driven model. Such model benefits from the learning process in an intelligent agents. The agents learns from observed data to provie a predictive model and help engineers for decision making.
- **Provide benchmark meta model**
  Create a meta model validated by experiments. Such model is utilized for real-world pipeline exposed to certain damages at different locations.
- **Utilize probabilistic predictive model**
  Create an intelligent agent that learns from data and provide a predictive model considering the uncertainty corresponding to the problem
- **Analyse sensed data**
  Creating a library of features (covariates) that is fed to the agent for learning purposes.
- **Damage detection and localization**
  Utilizing the previously constructed agent for damage detection and localization of the pipeline
- **Provide a guideline for decision making**
  Providing a suggestions and recommendations for engineers on how to use the agent's results for decsion making corresponding to damage detection and localization

1.3 Thesis structure

Here the goals and specifications of this study are provided in the form of stages that helps the reader to follow the whole manuscript.

Stage 1: Three-meter pipe modal analysis The modal test was is carried out on a 3m piece of a two-inch pipe used in inland lines to obtain the natural frequencies. This experiment is conducted at Sharif University of Technology. Although the design is more focused on interurban lines, the

choice of two-inch pipes has a particular reason: In vibrations new modes appear with increasing the length to diameter ratio; in fact, the larger the length/diameter ratio in simulation leads to the appearance of more natural vibrations appearing in vibrational behavior and the results are more realistic. Due to limited financial resources, it is not possible to test pipes longer than 3 meters. With a constant maximum length, the only way to have the most accurate laboratory results is to use a smaller diameter. It should be noted that the goal of the first two phases of this project is to validate (a) the finite element (b) the proposed software to implement this model, and (c) the type of proposed element and the test model results are not directly used in fault detection; so the use of a two-inch pipe, even if the goal is to detect the faults in longer pipes is justified. At the end of the first stage, the results are compared with the simulation results obtained from the second stage which confirms the validity of the model developed in the second stage.

Stage 2: Model At this stage, the pipe is modeled in Ansys software and under free- free conditions, similar to the experimental conditions, with the finite element method under vibration analysis. This stage includes increasing pipe length to 50 m and constraining the pipe to the actual state without considering the effect of soil and welding in the simulation space, determining the location of application of the excitation force so that its effect can be observed over a significant period, obtaining the maximum range of excitation such that an impact function would not cause damage to the pipe, creating 15 holes with a diameter of 2 mm in different lengths of the pipe and applying a force to the perforated and healthy models for 0.1 seconds and recording the specified signals.

Stage 3: Damage Library and features In this stage the numerical procedure to construct a FEM model is provided. The main objective of this section is to create new damage scenarios based on previous calibrated model. Finally, that stage aims to extract a library of (statistical) features. Basically, the attributes (features) of the sensed data are used to construct a library for supervised learning purposes.

Stage 4: Intelligent predictive model At this stage we use Gaussian Process Regression Method (GPRM) in order to construct a probabilistic predictive model. The model is updatable and used to address the concerns with respect to damage detection and localization. The details of GPRM are explained in Chapter.5.

Chapter 2

Literature review

Diagnosis of gas pipeline fault plays an important role in gas transportation both in terms of safety and economics. This chapter is dedicated to examine different faults and corresponding causes and structural health monitoring.

2.1 Methods of health monitoring for pipeline

Generally, there are two distinc approaches to detect faults along a pipeline as follows.

First approach In this approach the entire pipe is examined for fault detection, where the diagnostic device must be moved across the entire pipe or it is installed along the pipe. These include the use of optical or ultrasound sensors to find leakage [40]. Other examples include injecting flammable chemicals and using a flame detector along the pipeline [17] and simultaneous use of electromagnetic sources and detector sensors [14]. Moreover, another way is to use a special robot "pig" moving along the pipe, Fig. 1. These robots are commonly used to detect faults on the pipe, such as corrosion, cracking, and so on. In most cases, this expensive device is used to inspect gas pipelines that are not underground and it requires that the pipe would be out of service [21]. Another example for this category is the installation of fiber optic along the entire pipe [38]. It is worth noting that all of the aforementioned methods are demanding with respect to time and cost.

Second approach In this approach for gas pipelines' fault detection it is necessary to measure some variables in the limited points of the pipeline. Generally, in two ways, this approach can be addressed: (1) fault detection based on monitoring the changes in fluid characteristics (e.g. flow rate and pressure) [13] and [16] and 2) fault detection using ultrasound waves [35] as depicted in Fig. 2. Former method lies on the solution of nonlinear equations that describe the flow dynamics through linearization [36] or the discretization of nonlinear equations [12] - the rate or pressure is an indicator of presence/absence of fault. Nevertheless, this method suffers from the errors caused by the complex natural gas dynamics and the inherent uncertainty in the parameters of the governing

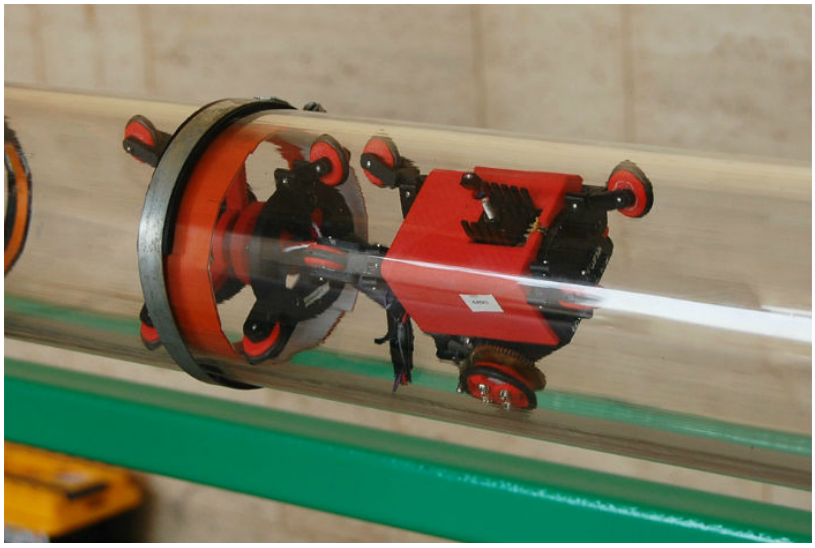

Figure 1: An example of Pig robot for leak detection. Source: *https://sourceable.net*

equations. In contrast, ultrasound waves have been successfully used to detect gas pipeline leaks[1]. The main shortcomings of this method are the scope of operating range (10 meters), high cost, and the utilization of generators and ultrasonic wave detectors.

2.2 Causes of gas leakage in pipelines and leak detection methods

In this section, a brief discussion of the main causes of gas leakage in pipelines Fig. 3 as well as suplementery leak detection methods are provided.

2.2.1 Causes of gas leakage

Gas leakage occurs due to various causes. In what follows, the list of primary causes are listed. It should be pointed out that in many practical cases, engineers face a combination of causes, which increase the complexity of the detection.

a) **Corrosion** is possible due to imperfections in the insulation or in the absence of correct cathodic protection of the outer surfaces of the pipes. Although corrosion mainly occurs at the outer surfaces of the pipes, internal corrosions still can happen either in pipes or in operating containers due to corrosive compounds such as hydrogen sulfide and water.

b) **Internal wear** phenomenon occurs due to the presence of impurities associated with gas. Typically, the amount of wear is higher in the bending of the pipes or in pressure relief valves due to increased gas velocity.

[1]This procedure works based on the speed of the sound in continuum. The reader is reffered to *http://www.uesystems.com/* for detailed explanation of technology and its application.

Figure 2: An example of ultra sound device for leak detection. Source: *https://www.engineering.com*

c) **External factors** such as mechanical shock, contact with power cord or induction currents.

d) Defect(s) in the **metallurgical structure** of the pipe, fittings, valves and other accessories.

e) Defect(s) in the **implementation and installation** of valves and other flange and thread connections.

f) Flaw in the **welding** of pipes and **fittings**.

2.2.2 Leak detection methods

In this section detailed methods for leak detection is discribed and suggestion for optimum monitoring is provided. The detection methods are divided in two categories, namely external and buried as well as special detection for global transmission lines [12].

Leak detection in external pipes and fittings In order to prevent potential leakage in the pipe and fittings, the following measures should be taken regularly.

a) The path of the gas pipelines should be checked objectively and be under constant supervision according to the smell and sound.

b) All paths are closely monitored and closure welds, connections, and valves should be detected by foam every six months. It is notable that in facilities leak detection is done by gas meters.

c) All external gas pipeline lines are checked by closing the valves and separating different parts from each other in terms of pressure drop every two years.

Figure 3: An exampl of surface gas pipeline. Source: *https://www.thomasindcoatings.com*

Leak detection in buried pipes In order to prevent potential leakage in the pipe and fittings, the following measures should be taken regularly.

a) Checking the pipeline path: The path should be inspected periodically and as soon as the smell of gas or leakage symptoms are detected, more detailed examination should be done and the detected leakage should be removed.

b) Periodic leak detection: In this type of leak detection which is recommended in short pipelines such as underground gas pipes in factories and commercial consumers, all consumer valves are closed and the sealed fuser inside the network is carefully measured and tested for a few hours and if there is a pressure drop, the leakage should be located.

Leak detection of global gas transmission lines Global gas pipelines and gas distribution networks are detected by Flame Ionization Detector (FID) leak detection devices as shown in Fig. 4. They are very sensitive devices that are used to detect the gas leakage in buried pipes and there are several types of them. "FID" is a leak detector used by the National Gas Company, which operates on the basis of hydrogen flame ionization in which there is a hydrogen fuel combustion chamber containing hydrogen fuel that absorbs the air from the environment and once the smallest amounts of gas hydrocarbons enter the chamber, the ionization of the flame environment changes and the result is seen as visual and audio signals. The amount of gas leakage will eventually reach the ground so these devices will be able to detect it with high precision. The accuracy of leakage detectors is as high as one per million PPMs so that when used in cities or industrial sites, they must be calibrated to the hydrocarbons in the air.

2.2.3 Leakage classification

Leakage may occur in one of the following ways that should be treated as a dangerous phenomenon in any case:

a) Gradual leakage from the gas piping system or gas appliances which that is gently accumulated

Figure 4: FID gas detector device. Source: *http://gasandoil.com.au/*

and flammable when released in the open space. This state usually occurs in cases that are insufficiently ventilated.

b) A sudden leakage with high volume from the piping or gas appliances that occurs due to fracture, mechanical failure or negligence occupies the space in short time and is ready to explode. In the case of the first type of leakage, the presence of steady gas leakage can be very useful and with the aid of these devices, it can be detected and stopped before the amount of gas in the environment reaches a dangerous level. In the case of leakage in the second group, prompt actions should be implemented and the slightest negligence in these cases will result in irreparable damage.

2.2.4 Gas detectors

Gas detectors are devices that detect the presence of gas at low concentrations below the explosive limit and alarm. These devices have a platinum catalytic filament that can provide the possibility of combining gas and oxygen at concentrations below the explosion limit and change filament temperature based on gas concentration and then the resulting flow will be specified as change in resistance in a circuit and then in the form of analogue or digital visual and sound signals on the screen. These devices are functionally divided into two groups, namely manual and portable gas detectors and fixed gas detectors

Manual and portable gas detectors These type of manual gas detectors are portable in the direction of the gas pipe or the desired areas and in the event of a collision or any leakage. Due to the sensitivity of the device, it alarms the presence of gas and its amount. The purpose of these gas-meters is as follows.

a) Supervising the external pipes in terms of possible leakage.

b) Measuring and ensuring the absence of gas prior to welding on pipelines or operational containers or any other hot operation that takes place inside the gas installation.

c) To ensure the absence of gas in enclosed areas such as ponds, reservoirs, engine room... before entering them

Fixed gas detectors These types of gas detectors that are used to detect gas permanently are installed in industrial environments, halls, or workshops and are divided into two categories, namely individual and network gas detectors.

a) Individual gas detectors: These types of gas detectors operate independently and most of their applications are non-industrial and are used in residential units, engine rooms, buildings, water utilities, etc. Fixed gas detectors are installed for natural gas at high level and low-level liquid gas.

b) Network gas detectors: This system consists of a central control center, and gas detectors installed at various locations. Their major use is in utility halls, warehouses or turbine chambers. The main display of the system which is connected to all sensing devices is located in the control room or fire department, and sensors are installed at points of installation or warehouse with the probability of gas leaks. Each sensor sends the signal to the control room as soon as it detects the gas to inform the user about the location and degree of leakage.

2.3 Summary

As it is investigated, there are different approaches towards monitoring pipelines. Utilization of the methods depends on the location and type of the pipes considering the cost and time. Again, it should be pointed out that there is no unique approach for monitoring and in many situations, engineers need to use a combination of methods to address the monitoring issues. As it is seen, there are no intelligent agents that are used frequently to monitor the pipelines. Pipelines are usually very long and monitoring of them can result in high cost and time demands. In addition, with the case of using robots, one particular issue is that although it uses intelligent agent, the time to monitor the pipeline is too long as well as expensive. On the other hand, using ultra sound seems promising because of the speed of waves in continuum. Therefore, introducing an intelligent agent that is trained for deferent fault scenarios based on ultra sound tests can be promising in reducing the downtime, cost, and time demands, significantly. In the next chapters we first introduce the pipeline under consideration and then we introduce a new Artificial Intelligence (AI) technique for damage detection and localization.

Chapter 3

Experiment and frequency domain analysis

This chapter introduces the modeling stages of the project and each one will be described in the subsequent sections. This includes experiment setup explanation, numerical simulation anad validation, and system response due to two different faults.

3.1 Pipeline experiment setup

The metal pipe is considered isotropic, and the free-free boundary conditions conditions are provided on both sides of the pipe. In addition, relatively soft spring is used to suspend the pipe so that the added frequency to the system (rigid frequency)[1] is negligible. The suspension system should be such that it does not affect the shape of the modes. In order to overcome this issue, it is suggested that the suspension system is installed such that it is perpendicular to the excitation directions. According to the results obtained by the finite element method (Ansys software), the movement of each of the four shapes of the first mode is perpendicular to the pipe axis and the suspension system is perpendicular to the direction of excitation. The highest rigid mode frequency should be less than 20% of the first non-rigid mode frequency (natural pipe) [7]. Therefore, the length of the rope used to hang the pipe should be such that the pendulum frequency of the structure is less than 20% of the first frequency obtained by the finite element method (Ansys software). As such, one can obtain the minimum length of the rope as follows.

$$\begin{aligned} 0.2\omega_1 &= \sqrt{\frac{g}{l_{min}}} \\ \Rightarrow l_{min} &= \frac{g}{0.04 \times \omega_1^2} \\ &= 0.152\ m, \end{aligned} \tag{1}$$

[1] It is worth noting that when the pipe is not fixed (anchored) in space domain, vibration modes can indicate rigid body motion in which the pipe can experience free translation or rotation without facing significant elastic fields.

where $g = 9.8\ m/sec^2$, $\omega_1 = 40.101\ Hz$, and l_{min} are gravitational acceleration, natural frequency of the pipe specimen, and the minimum length of the rope, respectively. The minimum acceptable rope length is about 15 cm. However, the length of the rope is selected about 1 meter in this experiment. As such, the rigid frequency of the system will be much lower than the first non-rigid frequency of the pipe. In addition, the location of the ropes is selected on the nodes corresponding to the first mode. Fig. 5 showls the experiment setup.

Figure 5: Experiment setup

3.1.1 Hammer test

In this experiment, the pipe is divided into 10 equal parts. An accelerometer is connected to the pipe and the impact is applied with a hammer (source of shock energy) to each 11 points (including pipe's ends). This work is repeated 5 times at each point so that the results are averaged to eliminate the possible noises. In this test, the data aqusition system is used to transfer and store data. It should be noted that it is attempted to allow the pipe to be perpendicular to the suspension system as much as possible. Failure to comply with this issue is the source of the error itself.

Sensor's mass Since the weight of the sensor used in this test is less than the total weight of the structure (about 5 g), the sensor weight does not affect the results.

Data acquisition The signal conditioning and signal generator parts are not detached from hardware and are embedded in NI 4310. Due to the use of an accelerometer in this test, a wire from the accelerometers is connected to the NI 4310. The hammer is also used in this experiment. The applied sensors and hammers are PCB type and anti-noise cables are used. Fig. 6 illustrates the framework used to perform the test.

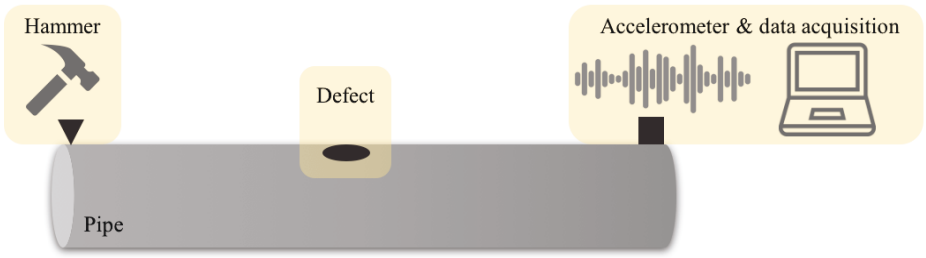

Figure 6: Schematic experiment setup.

3.1.2 Data processing and mode shapes

After installing the pipe and connecting the sensor and the hammer to the device, ModalVIEW software is utilized to perform the primary results related to vibration characteristics. It is worth noting that the type of sensors and their accuracy, location of the sensors, the impact location and the number of test steps must be defined prior to the analysis. According to Finite Element Method (FEM) results, the first four natural frequencies of the pipe are less than 1000 Hz, so the frequency range of the input power is considered to be 2048 Hz. Averaging is used to minimize the error caused by the noise as the experiment is carried out repetitively. Table. 1 shows the results from the experiment that will be used as the benchmark for FEM model calibration.

Table 1: Mode shape frequency results in proposed experiment

Mode Number	Frequency, Hz.
1	40.02
2	110.95
3	216.55
4	365.18

The first four mode shapes of the pipe as a function of its length L, $\bar{w}(x)$, is shown in Fig. 7. Interestingly, following [5], normalized mode shapes $\hat{\bar{w}}(\hat{x})$ can be expressed as

$$\hat{\bar{w}}\left(\hat{x}\right) = \sinh\left(k_n\hat{x}\right) + \sin\left(k_n\hat{x}\right) + \frac{\sin\left(k_nL\right) - \sinh\left(k_nL\right)}{\cosh\left(k_nL\right) - \cos\left(k_nL\right)}\left[\cosh\left(k_n\hat{x}\right) + \cos\left(k_n\hat{x}\right)\right] \quad , \hat{x} = \frac{x}{L} \in [0,1]\,, \tag{2}$$

where k_nL can be obtained by solving the following non-linear characteristics equation.

$$\cosh(k_nL)\cos(k_nL) = 1, \tag{3}$$

in which k_n is the wavenumber of nth vibrational mode expressed obtained from following equation with respect to the corresponding angular frequency ω_n.

$$\omega_n^2 = \frac{EI}{\rho A}k_n^4. \tag{4}$$

In Eq.4, E, I, ρ, and A are the Young Modulus, second moment of area of the cross section, density and cross section of the pipe, respectively[1]. The importance of having prior knowledge of the mode

[1]The mode shape function can be obtained by solving the Bernoulli–Euler beam differential equation,

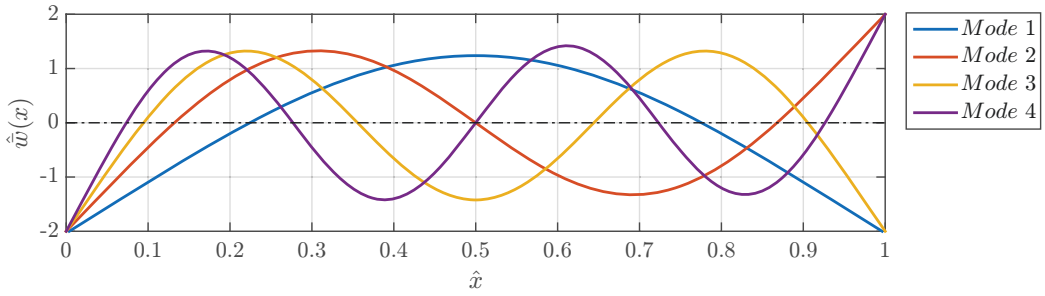

Figure 7: Normalized mode shapes of free-free pipe.

shapes helps us to determine the position of the hanging ropes. As the first mode plays the important role in vibration characteristics of the pipe. It is estimated that the ropes should be located at the distance of 20% of the length of the pipe from its both ends. Consequently, the effect of the hanging rope is negligible as it is located on the zero magnitudes points of the vibration corresponding to the first mode.

3.2 Frequency domain analysis and validation

In this section we utilize a numerical procedure using Finite Element (FE) method to construct a numerical model for frequency analysis purpose. The software ANSYS® is used to build such model to compare the performance of the model with the experiment. To this end, we compare the first four frequencies (mode shapes) of the modeled pipe with the experiment. Further more, we can extend the model and examine different defects along it. Here, we focus on the results induced from Discrete Fourier Transform (DFT), [33] and [32], and leave the detailed explanation of the modeling to the next chapter in which we build a library of damage scenarios. Before proceeding to the results obtained from numerical analysis, it is necessary to provide a brief introduction on Fourier transform and the application of FT in signal processing, particularly modal analysis.

3.2.1 Frequency domain analysis

In order to understand the behavior of the pipe under excitation, it is first necessary to prepare the resulted acceleration data. Then the obtained data are transformed from the time-domain to the frequency-domain. To this end, specific transformation is needed, namely Fourier transform. The Fourier transform named after the French mathematician Joseph Fourier is an integral transformation converting any time function such as $f(t)$ to a frequency function in the frequency field $F(\omega)$.As such, $F(\omega)$ is called the Fourier transform of the function. The transformation and inverse of are

$EI\partial^4 w(x,t)/\partial x^4 + \rho A \partial^2 w(x,t)/\partial t^2 = 0$, using the separtion of variables method for expressing the solution in spatial and temporal domain, $w(x,t) = \bar{w}(x)u(t)$

respectively shown the following in Eq

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt, \tag{5}$$
$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} F(\omega) e^{i\omega t} dt$$

Fourier transform and Fourier analysis are widely used in various topics of physics, including electronics, electromagnetic (especially in communication and telecommunication), acoustics, wave physics, and so on. [33]. In structural dynamics and mechanical vibrations the Fourier transformations are used to transform non-harmonic excitations to the harmonic components. Therefore, it enables us to solve the differential equation of the motion of the structure. Another application is in the analysis of telecommunication circuits and power circuits which are used to obtain the harmonics of a waveform [33].

Moreover, using Discrete Fourier transform (DFT), discrete functions and signals can be transformed from time-domain to frequency-domain (or from the location to the wave number field) or it can be used for the fast multiplication of polynomials and computer processing of signals. The complete Fourier transform means that it is possible to rebuild the initial signal from the transmitted signal, in other words no data is lost by applying Fourier transform and transformation is reversible [4]. Fast Fourier Transform (FFT) is the name of the algorithm for fast and efficient execution of direct and inverse Fourier transforms. FFT produces precisely those results that are directly derived by the definition of DFT. The only difference is that it is much faster. In practice, FFT decomposes a string of values into components with different frequencies. Calculating DFT for n points requires $\mathcal{O}(n^2)$ math operations, while FFT can lead to the same results in $\mathcal{O}(n)$ math operations [20]. This difference in speed can be very impressive, especially for the large data sets, where $n$is thousands or millions. Its improvement is approximately $n/\log^n$. This remarkable improvement has led to replacing DFT by FFT in a wide range of practical problems from digital signal processing to solving partial differential equations. It should be pointed out that Fourier transform algorithms deal with a very low error rate and a very small approximation error,this is mostly due to the floating points calculation.

3.2.2 Numerical model validation

Figs. 8-11 illustrate the first four mode shapes of the simulated pipeline using ANSYS® software. Two factors have to be verified: the modal shapes, and their corresponding frequencies. As it is seen, the shape of the modes are the same as of those determined in the previous section and shown in Fig. 7. In addition, Table. 2 shows the good agreement between the frequencies obtained from experiment and numerical experiment with the maximum error for higher modes. Again, the modeling procedure is discussed in the next chapter.

Table 2: Mode shape frequency results comparison between experiment and ANSYS®.

Mode Number	Frequency Hz. Experiment	ANSYS®
1	40.02	40.33
2	110.95	110.66
3	216.55	215.48
4	365.18	353.03

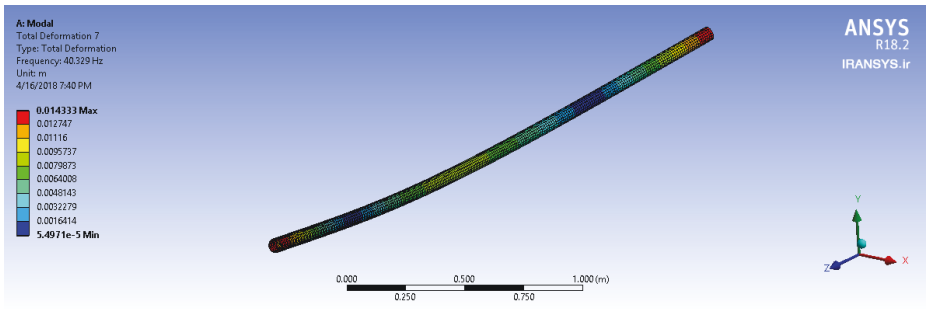

Figure 8: First mode shape of the modeld pipeline obtained from ANSYS®.

3.3 Summary

In this section first the experiment procedure is summarized and then it is attempted to build a numerical model and calibrate it with the experiment. To this end, we introduced the characteristic equations of the free-free end pipe to investigate the mode shapes. Afterwards, ANSYS® software is used to build a FEM model and the first four mode shapes were compared with the experimental results. The results show a good agreement between the numerical model and experimental one. Such calibration is crucial due to two main reasons. Firstly, it allows us to extend the model for real application (longer pipe with supports as shown in Fig. 3), and secondly we are able to introduce different types of damages for building a supervised learning model. In the next chapter, the framework for intelligent damage detection and location identification is provided. Also a detailed numerical procedure to build a library of damage scenarios is presented.

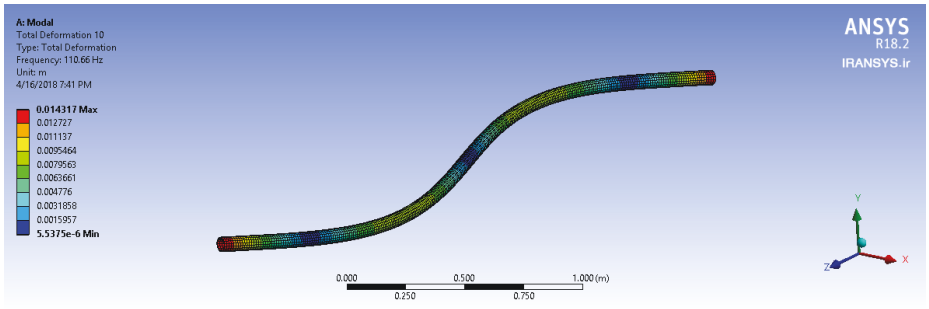

Figure 9: Second mode shape of the modeld pipeline obtained from ANSYS®.

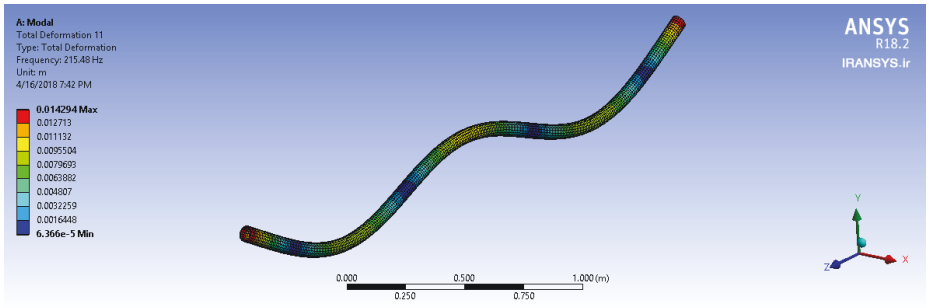

Figure 10: Third mode shape of the modeld pipeline obtained from ANSYS®.

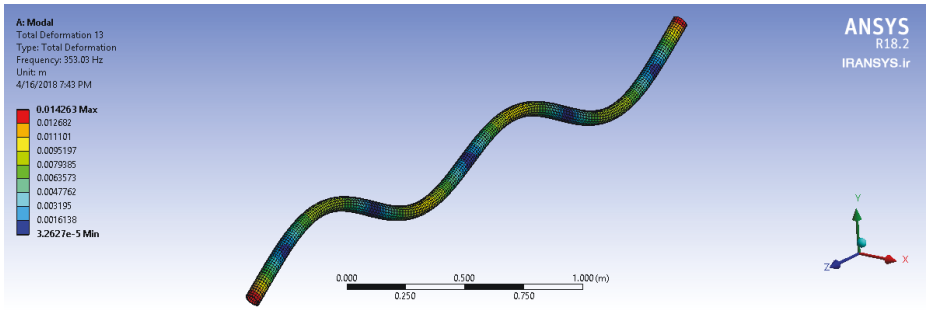

Figure 11: Fourth mode shape of the modeld pipeline obtained from ANSYS®.

Chapter 4

Intelligent damage detection detection and localization

In this chapter we introduce a framework for intelligent damage detection and localization of the surface pipeline due to induced defect. To this end, the organization of this chapter is as follows. In Section. 4.1 the intelligent framework is introduced and different components of it are explained. In the following sections, some preliminary stages are covered and the core of the intelligent component is left to be explained in the next chapter comprehensively. In Section. 4.2 the numerical procedure to construct a FEM model is provided. The main objective of this section is to create new damage scenarios based on a calibrated model. Finally, Section. 4.3 is dedicated to feature extraction, where information regarding some attributes (features) of the sensed data construct a library for supervised learning purposes.

4.1 Intelligent framework

In this section an intelligent framework for damage localization is introduced. Fig. 12 depicts the framework adopted and modified from [39]. The framework consists of two major procedures, namely meta model and SHM. In the meta model procedure the goal is to build an intelligent model from observed data and used the model for the other procedure. SHM procedure aims to detect damages and localize the damage based on the previously trained data-driven model. Each procedure includes certain sequential steps, which are described briefly in what follows.

Data acquisition This module is an indication of sensing different types of data under diverse conditions from various location of the structure.

Preprocessing In this module sensed data are preprocessed to improve the signal quality (e.g. averaging, filtering, and energy-normalizing). The reader is referred to the literature for a thorough discussion on different preprocessing approaches.

Feature extraction This module is dedicated to cast damage sensitive features from sensed data. It is suggested that a large library of features is needed as different features may be sensitive to different damages. Such features may be extracted by means of statistics, signal processing as well as machine learning techniques.

Gaussian Process (GP) At this step Gaussian Process is performed to create the intelligent agent for prediction purposes. Detailed explanation of the agent is provided in proceeding chapters.

Damage detection In this step, one may identify the existence of the damages by provided predivtive model.

Damage localization Having a predictive model along with the existence of the damages, one may determine the location of the previousy identified at this step.

Decision Having all extracted information from Data Processing and SHM Procedure, this module is dedicated to reconciling all results to allow engineers make a global decision for the system.

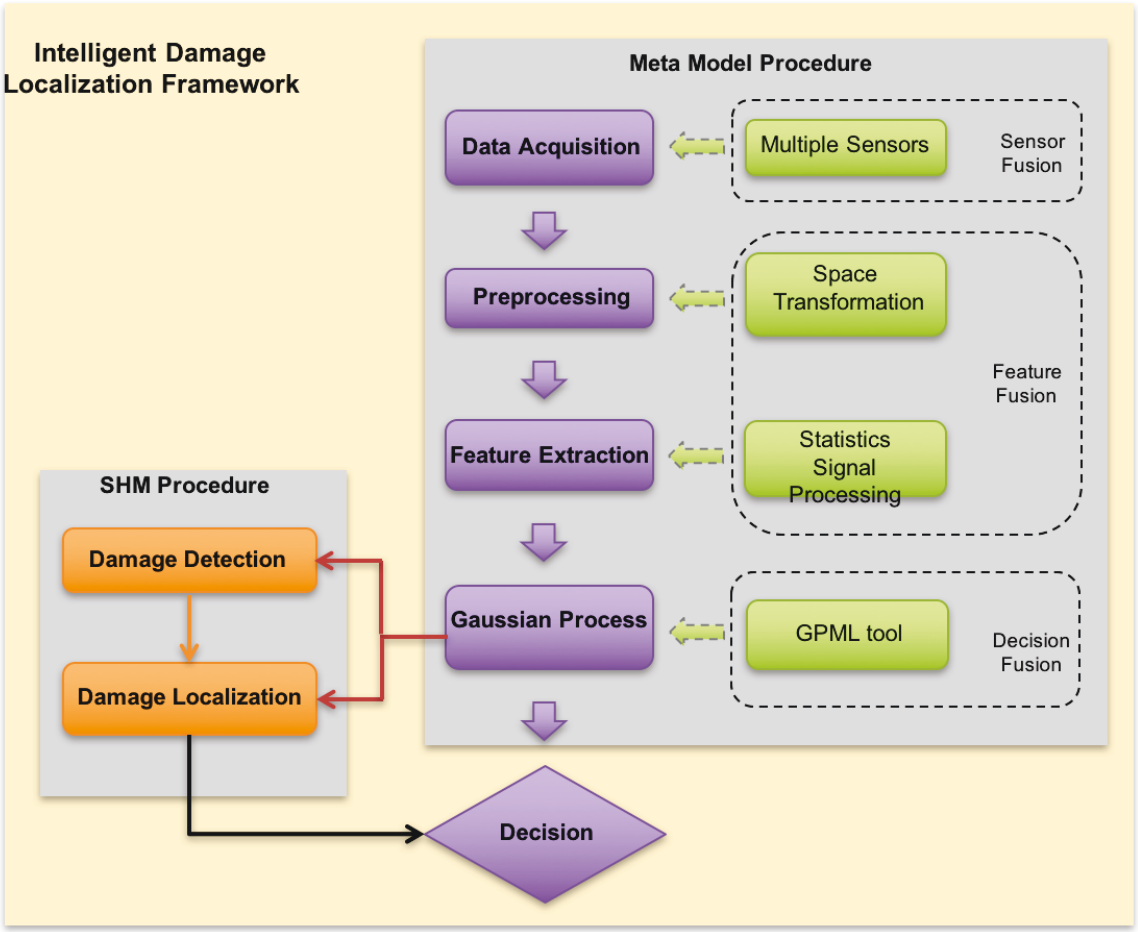

Figure 12: Intelligent damage localization framework, addopted and modified from [39].

4.2 Numerical Simulation

Utilizing FEM model to create new damage scenarios is the center of focus of this chapter. To this end, ANSYS® is used to address this issue. Here, a breif discussion regarding modeling and important considerations are provided.

4.2.1 Spatial characteristics

The schematic geometry of the pipe is shown in Fig. 13. As it is seen the pipe is modeled as a hollow pipe of length $50m$ with circular cross-section having the inner and outer radius of 25.14 and $30.15mm$, respectively. Every $5m$ along the pipe a simple support is modeled to indicate the support of the pipe in real applications. Finally, the Cartesian Coordinate System (C.C.S) is fixed at the one end of the pipe experiencing the impact loading.Additionally, another end of the pipe is the location of the sensor (here is the node that identified to read the induce acceleration), see Fig. 6.

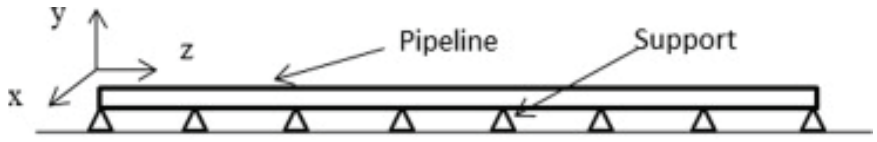

Figure 13: Schematic geometry of the modeled pipeline.

4.2.2 Mechanical characteristics

The pipe is considered as homogeneous isotropic material (i.e. steel) with the density, Young's modulus elasticity, and Poisson's ratio of $7861kg/m^3$, $207GPa$, and 0.3, respectively. It should be pointed out that the problem under consideration is elastic (linear) analysis, as such the pipe does not experience geometrical and material non-linearity such as large or plastic deformations, respectively.

4.2.3 Loading characteristics

Th pipe experiences the external impact loading of $F_{ext} = 300\ [N]$ and internal pressure of $P = 2068.43\ [kPa]$. The impact loading is calculated based on the maximum allowable stress design modified by endurance limit factors [28] in order to account for different real application factors such as loading, surface condition, temperature, and so on. Figs. 14 and 15 illustrate the location and the magnitude of the external and internal loading on the pipe. The impact loading is applied on the pipe as an impulse with the proposed magnitude between the first and second bearings. Such impulse can simulate the hammer test as mentioned before. In addition, the internal pressure is distributed along the inside of the pipe.

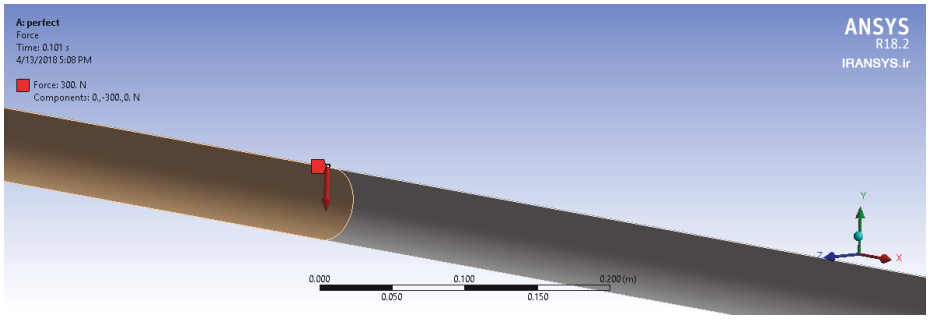

Figure 14: External loading on the modeled pipeline, ANSYS®.

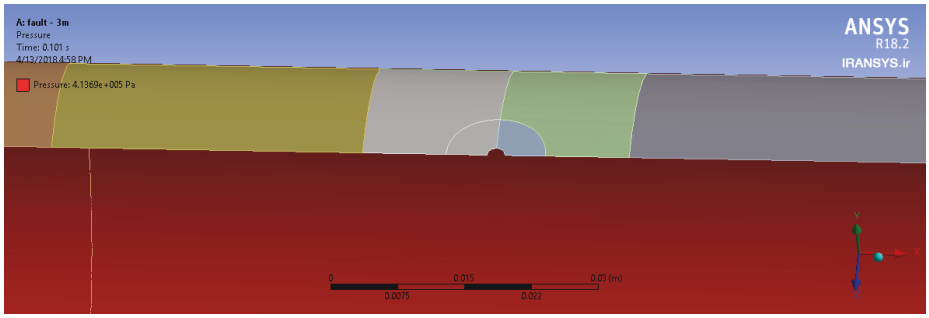

Figure 15: Internal pressure inside the modeled pipeline, ANSYS®.

4.2.4 Meshing

Meshing is the the crucial part of the simulation process. Although ANSYS® provides the ultimate solution, still there is a need for some meshing technique consideration specially when there are discontinuities such as cracks, holes, bi materials, and so on. The final meshing process resulted in 22509 nodes and 20000 elements. Faults are modeled as fifteen $2mm$ circular holes distributed every $3.25m$ on $+90^{o}$ position (top) of the cross section along the pipe. However, for those that are located on top of the bearings the distance is increased by $0.25m$ to prevent overlapping. Fig. 16 shows a typical meshing along the pipe where there is no hole. In the vicinity of holes the multi zone technique is used. As such, the area near the hole is divided in different zones and each zone has its own mesh grid and by closing to the hole the size of the element decreases to be able to capture the shape of fault as close as to the reality. Figs. 17 and 18 illustrate the multi zoning and corresponding mesh grid around a typical fault.

4.2.5 Dynamic simulation

The created model is exposed to dynamic loading (impact) and the vibration of a certain point(s) is collected for damage identification purposes. However, the reader should bear this in mind that defining correct time steps, where the loadings can be added or removed separately, is vital as at this point the numerical analyst tries to create the same condition as of the reality. In this manuscript, the total time of analysis is $t_{total} = 0.1\ Sec$ ($\Delta t = 5E-5\ Sec$) and it is divided in three sub-steps,

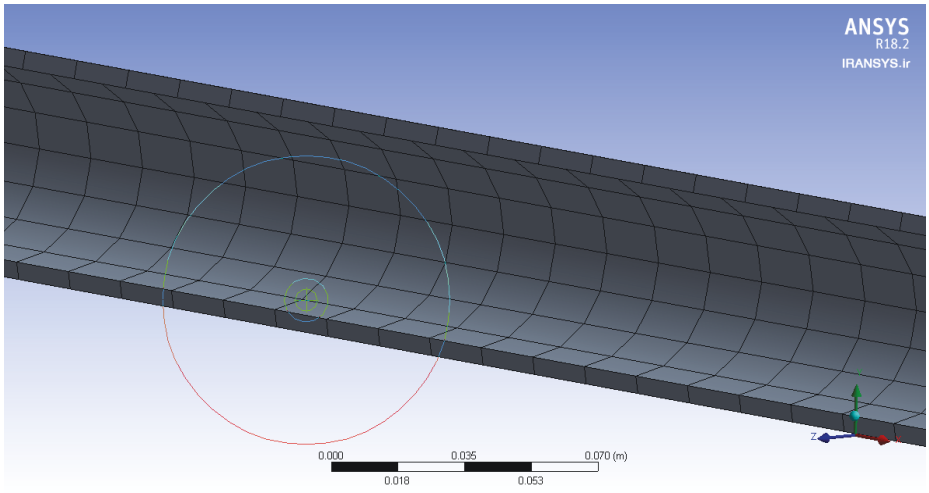

Figure 16: A typicsl mesh grid along the pipe, ANSYS®.

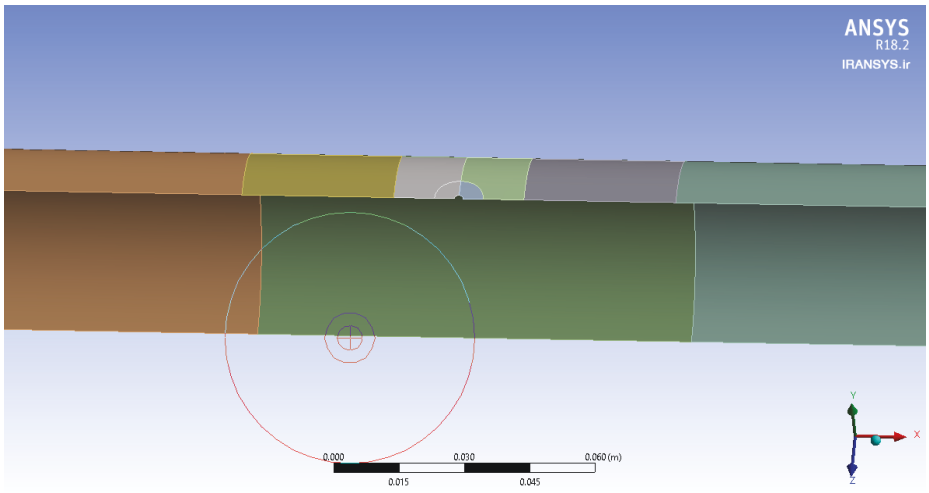

Figure 17: Multizoning in vicinity of the fault, ANSYS®.

where the loadings can be added/removed/maintained. Figs. 19 and 20 show[1] respectively the steps that the external and internal loading is applied to the pipe. As it is seen, the external loading is applied and removed in a small period of time in order to simulate the impact loading, while the internal pressure remains during the analysis.

4.2.6 Simulation results

The pipe is modeled with different one-point fault. Initially, the pipe is modeled without any fault indicating the intact (undamaged or healthy) structure.The undamaged model is used as the baseline to compare the results with. Afterwards, one-point hole as a fault is simulated at different

[1] As the figures are the outputs of ANSYS®, the axes are not labeled. The vertical axis is the magnitude (considering the loading direction) of the corresponding loading and the horizontal axis is the timestamps.

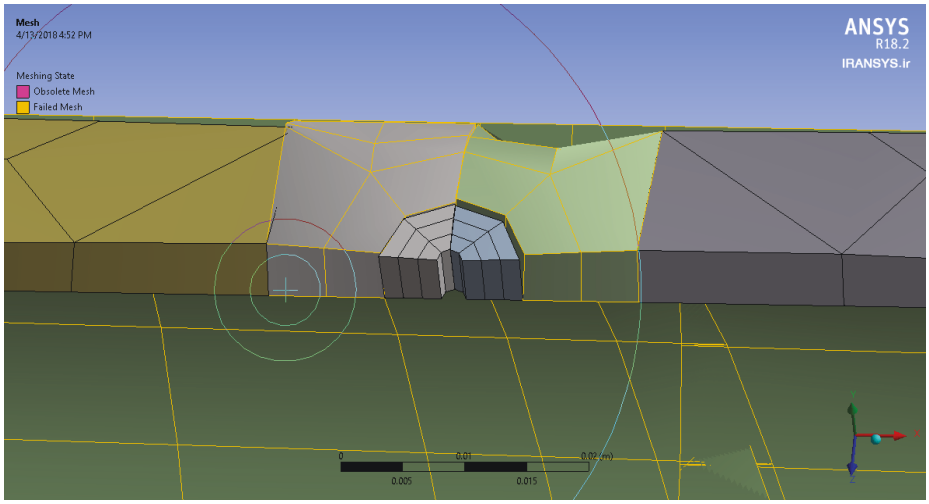

Figure 18: Induced mesh grid near the fault using multizoning technique, ANSYS®.

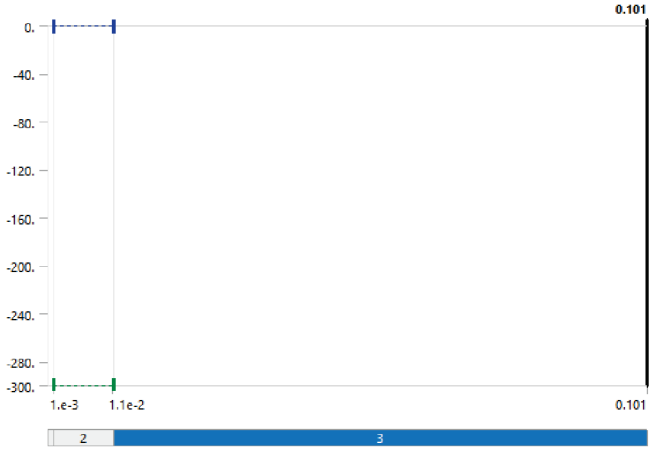

Figure 19: Load monitoring during daynamic analysis (external loading), ANSYS®.

location of the pipe. It should be pointed out that in each case (except the intact model) only one fault is introduced to the pipe. Fig. 21 illustrate two results corresponding to the intact pipe and the pipe exposed to a fault at 15.5 m along with it. As it is seen introducing fault changes the dynamic behavior of the pipe and therefore it can be used to identify the fault and it's location.The reader is referred to the Cahpter. A for a complete results of the responses. It should be noted that the difference between the response of the intact and damaged pipe are small. Thus, there is a need to extract damage-sensitive features from the data. Fig. 22 shows the absolute difference between the response of the pipe in two cases, intact and damaged. In next section a library of the features are extracted for machine learning purpose.

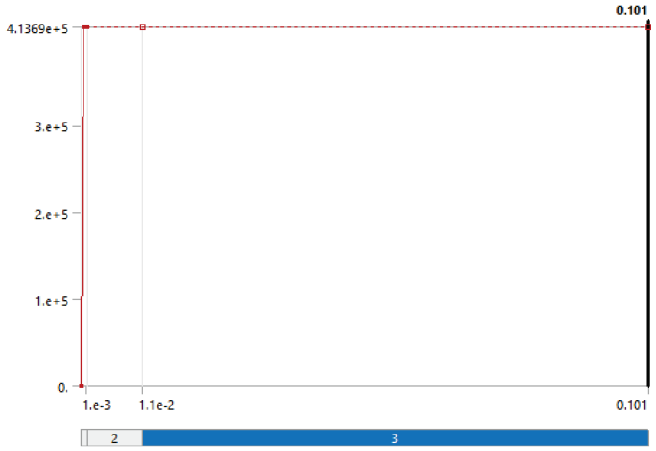

Figure 20: Load monitoring during daynamic analysis (internal loading), ANSYS®.

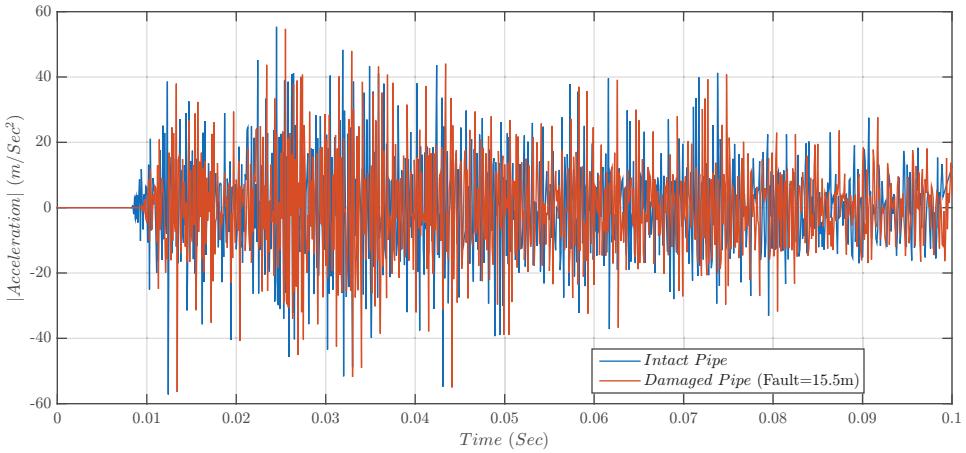

Figure 21: Acceleration response of the pipe in two cases (intact and damaged).

4.3 Damage library

4.4 Features

The performance of the algorithm is based on the effectiveness of the extracted features in action, it is difficult to find one universal feature that is sensitive to all types of damage, while in the presence of a variety of environmental changes is strong for all types of structures.Therefore, from this approach, one can extract many potentially suitable features to create a machine learning library and then use to automatically search feature, in this case the most appropriate candidate for channel specific features of the task. Figs.23-29 show the obtained feautures. Each feature shows the normalized value of the feature corresponding the location of the damage. In another words, the horizontal axis shows the pipeline location and each point is the value of the feature if there is a fault on that particular location. In addition, the term *Observation* is used to show the value we observe

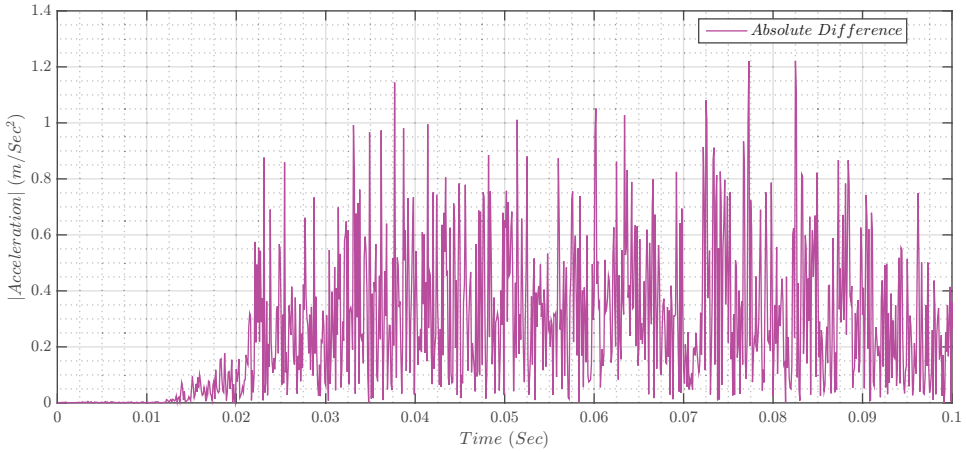

Figure 22: Abosulte response difference of the pipe in between two cases (intact and damaged).

(obtained) in feature-space due to the corresponding damage location. Moreover, It is worth noting that the term "Observation" is also not related to the location of the sensor, but as mentioned, it is corresponding the measured value in feature-space. Consequently, unless otherwise stated, such term is used to show the measured value in the feature-space, where the damage is introduced to the pipeline.

4.4.1 Energy Norm

Also known as the Euclidean norm, the L2-norm is a metric of the location of a distribution. For an n-sample discrete signal, the L2-norm is defined as the square root of the total energy of the signal:

$$||x||_2 = \sqrt{\sum_{i=1}^{n} x^2(i)} \tag{6}$$

4.4.2 Curve length

The curve length of a signal is useful for describing the signal complexity.A variation in curve length may be caused by changes in the modal amplitudes or locations of waves. The curve length of a discrete signal domain x(i) is defined by:

$$L = \sum_{i=2}^{n} |x(i) - x(i-1)| \tag{7}$$

4.4.3 Local Maximum Feature

The peaks of a complex signal indicate the arrival, reflection, or conversion of wave modes. Certain peaks are expected to be affected differently from others when damage is introduced. Local maxima (any sample larger than both of its neighbors) are searched over different signal domains. Features are constructed from the first, second, and third greatest peaks, and the peaks with the amplitude larger than 20 and 60% of the greatest peak.

Table 3: Damage location-covariates table.

ID	$peak \geq 60\%$	$peak \geq 20\%$	L2	CL	KU	COR	MeanPeak	Location
1	0.2621	0.1134	0.391	8.15E+03	183.2794	0.0201	0.0062	0
2	0.3106	0.1123	0.4501	1.77E+04	143.9647	0.9978	0.0114	3
3	0.3098	0.1221	0.4466	1.76E+04	146.3256	0.9981	0.0113	6
4	0.3112	0.123	0.4468	1.75E+04	147.3997	0.9979	0.0112	9
5	0.3171	0.1066	0.4517	1.75E+04	148.5618	0.9898	0.0115	12
6	0.3077	0.1109	0.4466	1.77E+04	143.3788	0.9979	0.0114	15.5
7	0.3132	0.1227	0.4498	1.77E+04	147.9885	0.9981	0.0111	18
8	0.3097	0.1227	0.4482	1.77E+04	144.5794	0.998	0.0116	21
9	0.3148	0.1235	0.4519	1.77E+04	148.2855	0.998	0.0112	24
10	0.3117	0.1234	0.4503	1.77E+04	144.9855	0.9981	0.0114	27
11	0.3070	0.1107	0.446	1.77E+04	143.0501	0.9979	0.0114	30.5
12	0.3131	0.1236	0.4509	1.77E+04	146.2843	0.9981	0.0116	33
13	0.3115	0.1232	0.4492	1.77E+04	146.1542	0.9981	0.0113	36
14	0.3203	0.1247	0.4559	1.77E+04	151.4946	0.9979	0.0114	39
15	0.3129	0.1233	0.4501	1.77E+04	147.0621	0.9981	0.0115	42
16	0.3066	0.1109	0.4458	1.76E+04	143.0101	0.9979	0.0114	45.5
17	0.2621	0.1134	0.391	8.15E+03	183.2794	0.0201	0.0062	50

4.4.4 Kurtosis

Kurtosis is generally used as a descriptor of the shape or peakedness of a distribution. Any mode conversion may change the shape of the peaks. Kurtosis of a discrete signal domain x(i) is defined as the standardized fourth central moment:

$$K = \frac{\frac{1}{n}\sum_{i=}^{n} |x(i) - \overline{x(i)}|^4}{(\frac{1}{n}\sum_{i=}^{n} |x(i) - \overline{x(i)}|^2)^2} \tag{8}$$

4.4.5 Correlation Coefficient

The Pearson product-moment correlation coefficient (PCC) is a common metric for similarity between two signals. The PCC of a discrete signal domain x(i) is: given by:

$$C = \frac{\sum_{i=}^{n} |x(i) - \overline{x(i)}||x_b(i) - \overline{x_b(i)}|}{(\sqrt{(\sum_{i=}^{n} |x(i) - \overline{x(i)}|^2)})\sqrt{(\sum_{i=}^{n} |x_b(i) - \overline{x_b(i)}|^2)}} \tag{9}$$

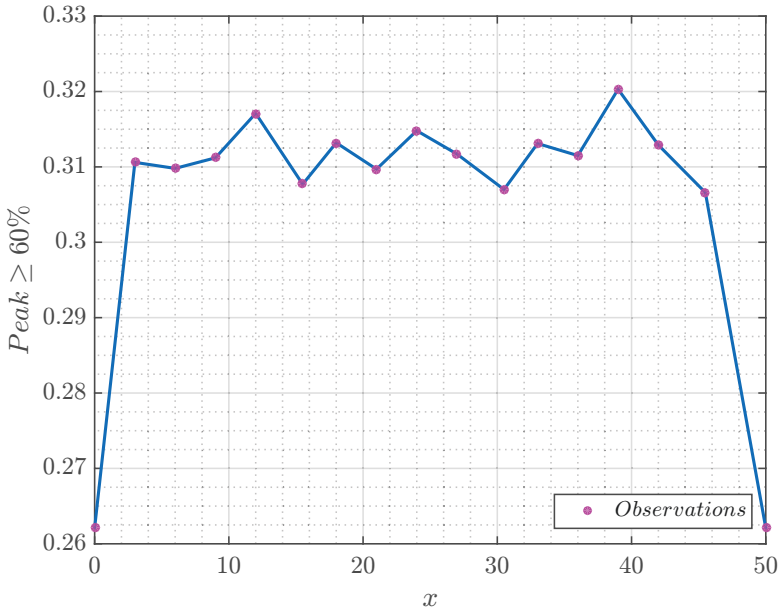

Figure 23: $peak \geq 60\%$

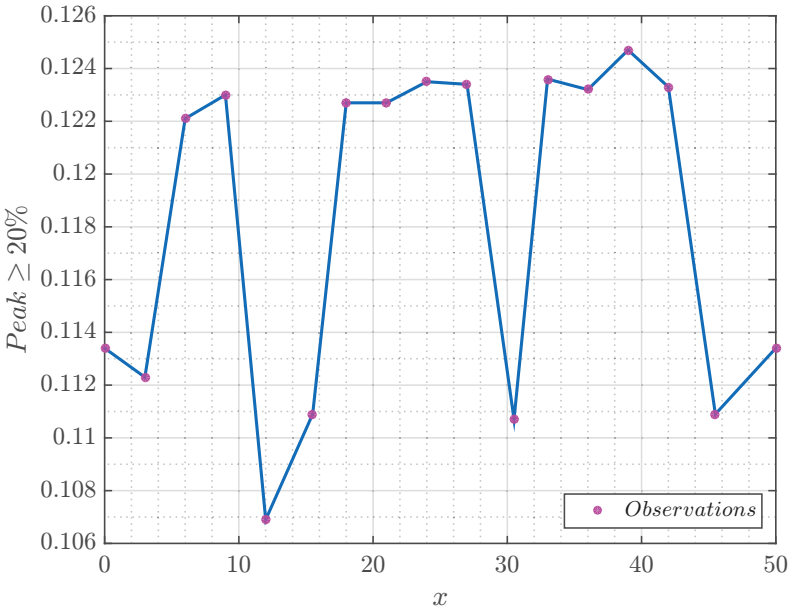

Figure 24: $peak \geq 20\%$

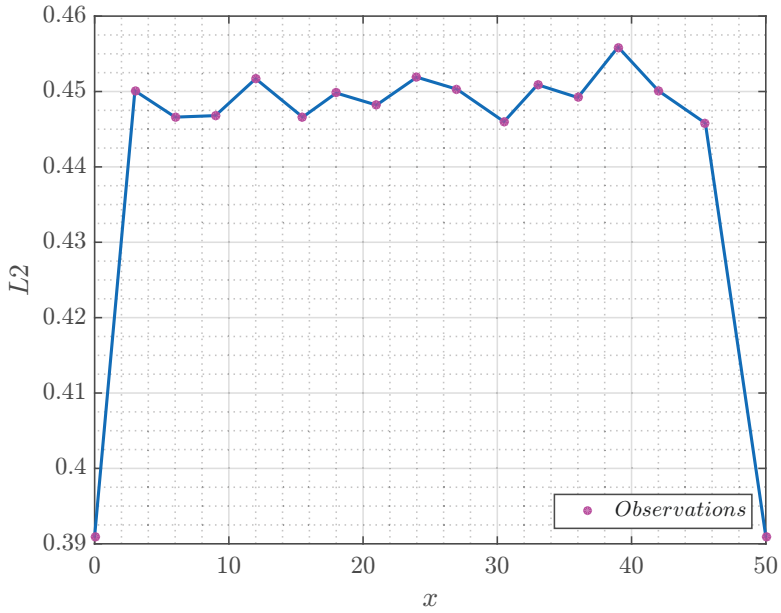

Figure 25: L2.

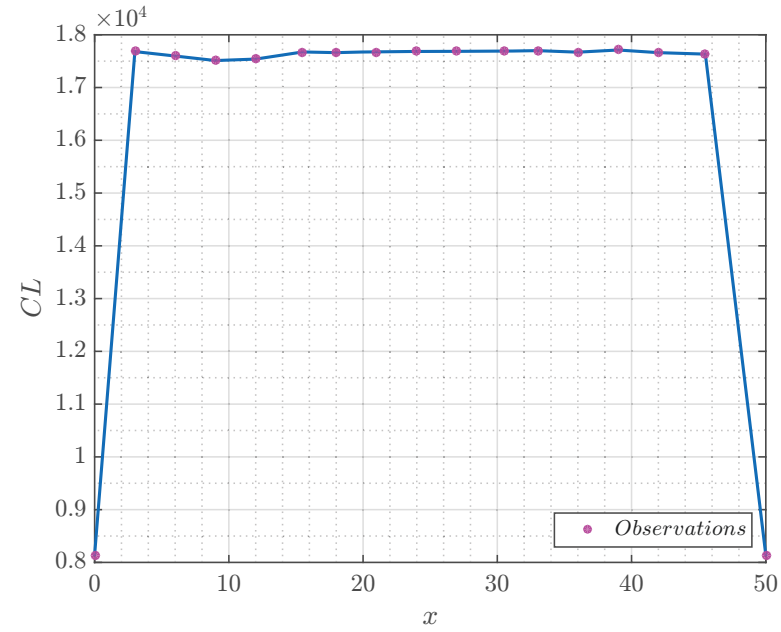

Figure 26: CL.

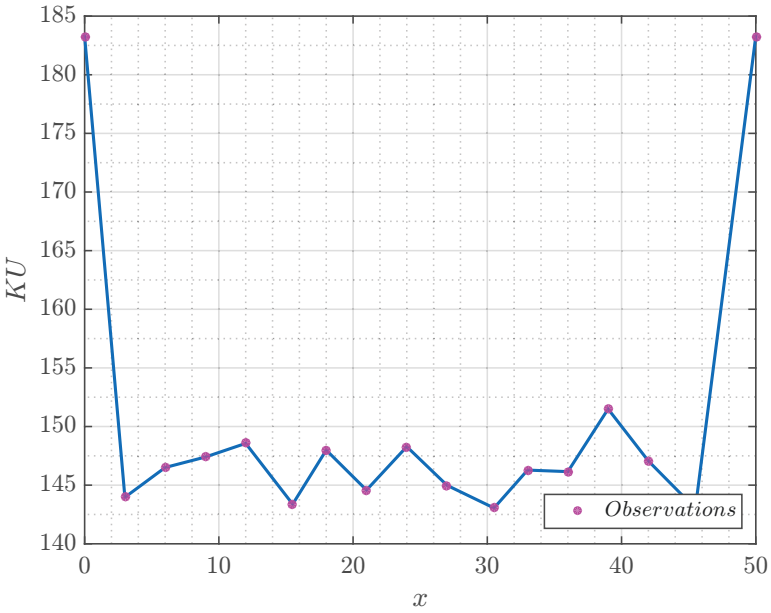

Figure 27: KU.

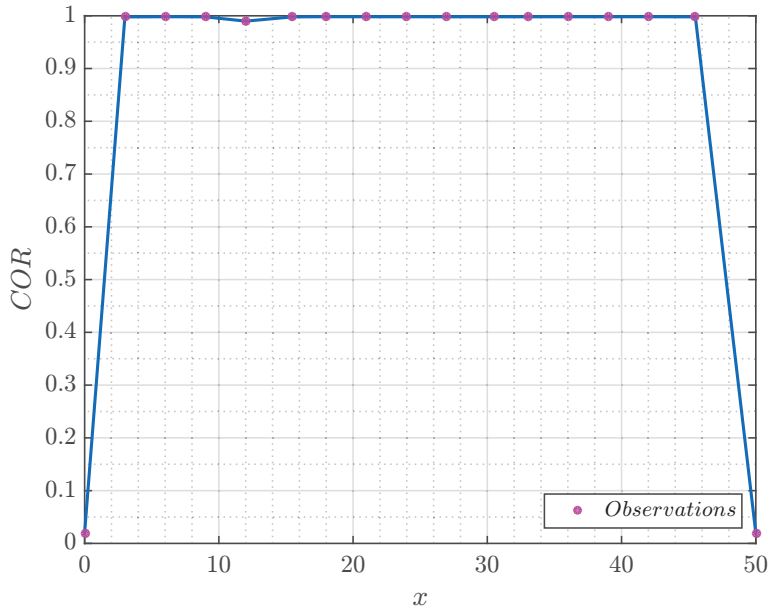

Figure 28: COR.

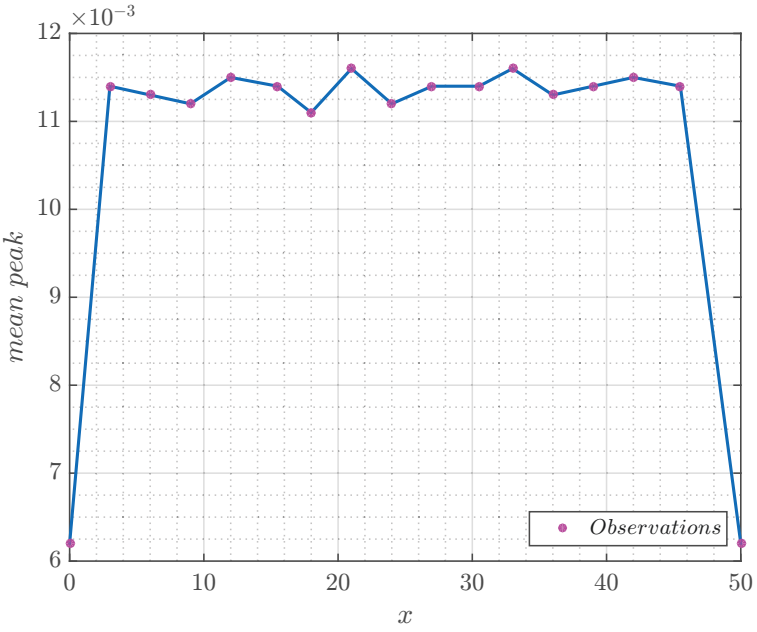

Figure 29: MeanPeak.

Chapter 5

Gaussian Process Regression Method

In this chapter first,the Gaussian Process (GP) for Machine Learning (ML) purposes is described and then the use of GP for damage location identification is investigated. This chapter is divided into four sections as follows. Section. 5.1 intends to answer the most important question, which is "*Why is it ought to use Gaussian Process in certain machine learning problems?*". Answering this question leads to presenting the advantages of using such methodology. In section. 5.2 the Gaussian Process Regression Method(GPRM) is explained and different kernel functions are introduced for the problem under consideration. In section. 5.3 updating procedure by means if Bayesian Theorem is explained in order to draw inference in light of new data. Section. 5.4 is dedicated casting available methods to learn parameters from data, and in section. 5.5 the problem of pipeline under various damages is modeled as a meta model and relevant considerations are discussed. Moreover, at the end of each section, the required parameters for th problem under consideration are identified.

5.1 Background

The aim of supervised learning is to learn input-output mapping from first-hand (observed) data. Generally, two approaches are available [22].The first approach deals with *restricted* class of functions to perform the mapping between input and output. Linear regression is an example of this approach in which the output is the linear combination of certain functions [23]. The problem with this approach is that we have to decide which types of functions have to be used. Subsequently, wrong choice of functions results in poor predictions. The second approach is to define a prior probability to every function and increase the probability for those which are more likely represent the target (final) function. However, the problem with this approach is the existence of infinite set of functions and lack of knowledge with respect to *preference* of each. The reader is referred to [30] for details. Gaussian Process (GP), which is the generalization of probability distribution, can solve the above-mentioned issues such that the probability distribution describes the random variable uncertainty, while the stochastic process deals with the characteristics of the proposed function [30]. In other words, choosing a function that is expressed in terms of probability distribution(s) can address the preference and restriction of the function at the same time. Another perspective that is needed to be

considered is related to the goals of *machine learning* and *statistics* and the type of problems they try to address. Generally, in statistics, the main goal is to understand the data by providing models that can express the relationship(s) and dependencies (*descriptive*), such as regression analysis, while the purpose of the machine learning is mainly to make accurate predictions (*predictive*), for instance, Neural Network.[1]. In contrast, GP cope with both school of thoughts by bridging the gap between both sciences. As such, GP provide a prediction tool from a finite set of empirical data as well as explaining the dependencies between the attributes (features) of problem under consideration [30]. Turing to the current problem in this manuscript, the pipeline exposed to a certain defect in various location and the dynamic system response is observed for each case separately. The objective is to propose a model that predicts the new system response due to never-seen-before defect and infers the location of such defect with the same model. In addition, due the epistemic uncertainty of the problem such model has to be able to provide such uncertainty whenever the observation is not available or there is an uncertainty during the observation.

5.2 Gaussian Process Regression

Gaussian process is the generalization of the Gaussian probability distribution and the Gaussian process Regression Method (GPRM) works by updating the knowledge about system responses in light of new observations. Given an observation from a system response one can write:

$$g_i = g(\boldsymbol{x}_i) + v_i, \quad \boldsymbol{x}_i = [x_1, x_2, \ldots, x_{\mathtt{X}}]_i^{\mathsf{T}}, \tag{10}$$

in which, g_i is the observation, $\boldsymbol{x}_i$ is the attribute vector, $g(.)$ is the the realization of the corresponding attribute, and v_i is the measurement error. The subscript i indicates the i^{th} observation from the pair-wised dataset $\mathcal{D} = \{(\boldsymbol{x}_i, g_i), \forall i = 1 : \mathtt{D}\}$, where $\mathtt{D}$ is the number of observations. In addition, the measurement errors are estimated as zero-mean normal distribution $v : V \sim \mathcal{N}(v; 0, \sigma_V^2)$, where the observation errors are independent (i.e. $V_i \perp\!\!\!\perp V_j, \forall i \neq j$). It should be pointed out that the problem under consideration is simulated by computer and therefore, the measurement error is equal to zero.(Sometimes input measurement errors occur when the geographical locations of observed data are not known exactly and there are noise in our experiment. Such sources of error are requiring variant methods for parameter estimation. Gaussian process models do not straightforwardly extend to incorporate input measurement error, and simply ignoring noise in the input space can lead to poor performance for prediction. Reviewing and extending existing theory on prediction and estimation in the presence of location errors and noises can show that ignoring location errors may lead to Kriging that is not self-efficient. Also, Markov Chain Monte Carlo (MCMC) approach using the Hybrid Monte Carlo algorithm that obtains optimal (minimum MSE) predictions, and discuss situations that lead to multimodality of the target distribution.) Hence, having the same notations,

[1]The reader should pay attention that here the distinction is between the classical statistics and relatively new machine learning paradigms. Apart from predictive and descriptive modeling, there is a third class of modeling known as explanatory that utilize data and statistical learning (or data mining) approaches in order to explain the data and their causal relationship [31]

the GP can be rewritten as

$$g_i = g(\boldsymbol{x}_i), \quad \boldsymbol{x}_i = [x_1, x_2, \ldots, x_{\mathtt{X}}]_i^{\mathsf{T}}, \tag{11}$$

As mentioned, in GP the system response is assumed to be a realization from a Gaussian process, so $g(\boldsymbol{x}) : \mathbf{G}(\boldsymbol{x}) \sim \mathcal{N}(g(\boldsymbol{x}); \mathbf{m_G}, \mathbf{\Sigma_G})$, where $\mathbf{m_G}$ and $\mathbf{\Sigma_G}$ are respectively the prior mean vector $[\mathbf{m_G}]_{ij} = m_{\mathbf{G}}(\boldsymbol{x}_i)$ and covariance matrix defined as:

$$[\mathbf{\Sigma_G}]_{ij} = \rho(\boldsymbol{x}_i, \boldsymbol{x}_j).\sigma_G(\boldsymbol{x}_i).\sigma_G(\boldsymbol{x}_j). \tag{12}$$

In Eq.12, σ_G is the standard deviation of the gaussian process and $\rho(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is known as correlation function[2] indicating the dependencies between two sets of random variables, $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. The mean vector and covariance matrix are all together represents our prior belief about the system before presenting the observation. In addition, the choice of correlation function determines the kinds of function that we expect to obtain after introducing the empirical data (observation). So, the choice of correlation function is crucial to obtain desirable results. Refer to [30] there are well-studied correlation functions, in which this study focus on two of them namely, *Squared Exponential* and *Periodic* Class of Covariance Functions. In what follows a thorough discussion is provided to compare and justify their utilization in pipeline problem.

5.2.1 Correlation function

Before proceeding to treat the problem under consideration it is necessary to provide a comprehensive discussion on correlation functions and their effects on interpret data. Here, the focus is on two correlation functions, namely Squared Exponential and The Periodic Class of Covariance Functions. The formulation of both functions are as follows. For the sake of brevity and better illustrations, we use univariate functions rather than vector implementation. By understanding as such, one can readily extend the formulation for a multiple variables. In addition, in this study we examine the proposed features, separately. Using multiple variable is out of the scope of this study and is briefly explained in the next chapters.

Squared Exponential Covariance Function Squared Exponential (SE) covariance function can be written as:

$$\rho_{SE}(r_{i,j}) = \exp\left(-\frac{r_{i,j}^2}{2l^2}\right), \quad r_{i,j} = x_i - x_j, \tag{13}$$

in which, l is *characteristic length-scale*. As it will be discussed l plays an important role to determine the dependency between two random variables. Fig. 30 shows the correlation function for different values of l. As it is seen, by increasing l the dependency between the points increases. In another words, the effect of one point can spread far by increasing the characteristic length-scale. In the context of the current study, one may conclude that by increasing l the effect of the deffect can

[2] In statistics ρ is known as correlation factor, scalar quantity, indicating the linear relationship between two random variables. Here, such factor is in fact replaced by a specific function and as it will be seen, the parameters of this function are updated by seeing data.

be propagated far along the pipe. On the other hand, by decreasing l the effect of the defect is vanished rapidly (see Fig. 30 for $l = 0.5$). Estimating l is the objective learning procedure presented in Section. 5.4.

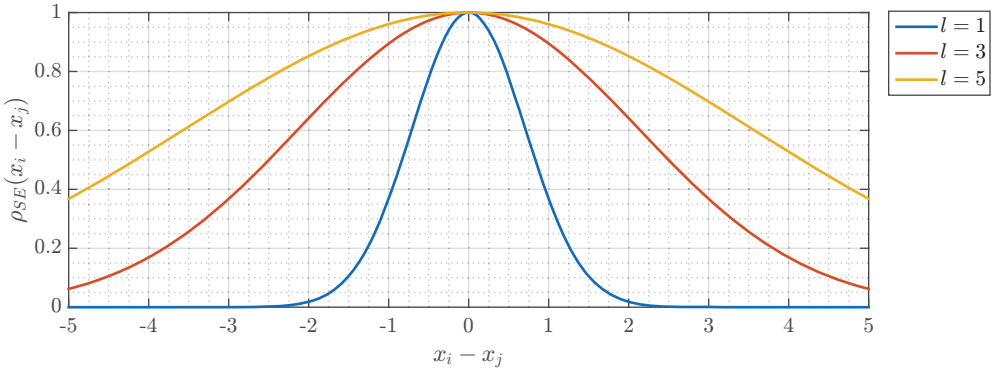

Figure 30: Squared Exponential covariance function with different characteristic length-scale.

Periodic Covariance Function Periodic (P) covariance function can be written as:

$$\rho_P(r_{i,j}) = \exp\left[-\frac{2}{l^2}\sin\left(\frac{\pi r_{i,j}}{p}\right)^2\right], \tag{14}$$

where, p is the period length. Fig. 31 illustrates the changes in the proposed covariance function with respect to l, while the periodic length $P = 3$. The same phenomena as previously seen for squared exponential covariance function; The greater the decrease in l, the sooner the effects vanish. In addition, Fig. 32 shows the role of the second parameter, p on the dependency. Briefly, periodic covariance function creates periodic dependencies over the domain. And, by decreasing p one can expect more periodic dependencies at different location of the problem domain (i.e. pipeline).

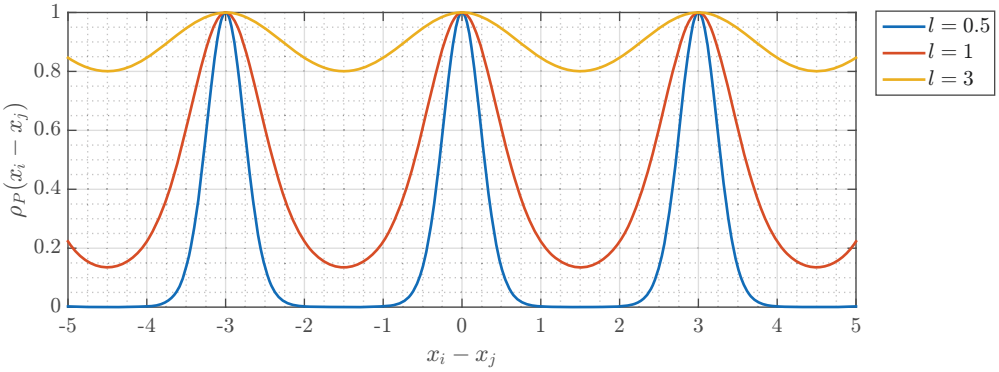

Figure 31: Periodic covariance function with $p = 3$ and different characteristic length-scale.

Comparison between two aforementioned covariance functions reveals the fact that the periodic covariance function is in fact an extension of the squared exponential covariance function that

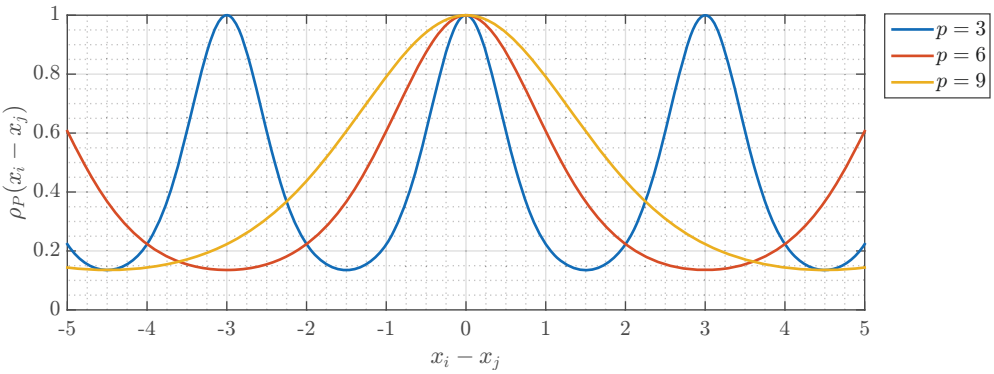

Figure 32: Periodic covariance function with $l = 1$ and different peridic length.

consists of periodic behavior. The pipeline under consideration along with its defect(s) represents a symmetric periodic behavior as shown in the figures by during numerical analysis. Therefore, the choice of square exponential and periodic covariance functions can addressed the learning procedure problem. It is worth noting that one may make new covariance functions from existing ones [30]. However, here we try to use them separately to investigate the effects of each of them for fast detection.

5.2.2 Feature selection

The characteristic length-scale l can be seen as *relevance* factor as introduced in [24]. Because the inverse of l in both Eqs. 13 and 14 indicates how relevant the selected feature is to the problem; The larger the l is, the less relevancy we except from the feature. In another words, utilizing such covariance functions can act as feature selection. As explained in [24], such parameterization with respect to l results in Automatic Relevance Determination (ADR). For instance, in Fig. 30 the yellow line, which corresponds to $l = 5$ indicates a dependency of all points along the domain. It means that for the choice of $l = 5$, the corresponding feature becomes irrelevant to the problem domain. As a general rule, if l is greater than two standard deviation of corresponding feature, the relevancy of such feature is negligible. It should be pointed out again that the choice of l is the objective of the learning procedure form data. Now, the question arises here is "*What types of covariance functions do perform better?*". The choice of covariance function, as mentioned, depends on the expected behavior from th analysis under consideration. There are approaches such as Bayesian model selection, cross validation, and automatic selection, which are beyond the scope of this study. The reader is referred to [37], [41], and [6], respectively, for detailed explanations.

5.3 Updating GP using noise-free observations

In this section, we introduce the approach for updating the Gaussian Process with the aid of exact observation obtained from numerical simulations (or experiments). Following , given Eq. 11 and pair-wised dataset $\mathcal{D} = \{(\boldsymbol{x}_i, g_i), \forall i = 1 : \mathtt{D}\}$, the objective of updating is to use exact observations (i.e.

noise-free observations) to update our knowledge about the posterior PDF $f\left(\boldsymbol{g}_*|\boldsymbol{X}_*,\mathcal{D}\right)$ knowing the prior one. As mentioned in section. 5.2 the system response is the realization of Gaussian Process, therefore, the posterior PDF with estimated mean vector $\mathbf{m}_*$ and covariance matrix $\boldsymbol{\Sigma}_*$ can be writen as:

$$f_{\mathbf{G}_*|\mathbf{x}_*,\mathcal{D}}\left(\mathbf{g}_*|\mathbf{x}_*,\mathcal{D}\right)=\mathcal{N}(\mathbf{g}_*;\mathbf{m}_{*|\mathcal{D}},\boldsymbol{\Sigma}_{*|\mathcal{D}}) \tag{15}$$

where,

$$\mathbf{m}_{*|\mathcal{D}}=\mathbf{m}_*+\boldsymbol{\Sigma}_{\mathbf{G}_*}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{G}}^{-1}(\mathbf{g}-\mathbf{m}_{\mathbf{G}}), \tag{16}$$

and,

$$\boldsymbol{\Sigma}_{*|\mathcal{D}}=\boldsymbol{\Sigma}_*-\boldsymbol{\Sigma}_{\mathbf{G}_*}^{\mathsf{T}}\boldsymbol{\Sigma}_{\mathbf{G}}^{-1}\boldsymbol{\Sigma}_{\mathbf{G}_*}. \tag{17}$$

It should be pointed out that in Eqs. 16 and 17, updating procedure is based on defining joint prior knowledge that we have for the observed and predicted locations, $\mathbf{x}$ and $\mathbf{x}_*$, respectively. The superscript $*$ is an indication of predicted values. As such, the joint prior knowledge for both observed and predicted system responses can be written as follows.

$$\begin{bmatrix}\mathbf{G}\\ \mathbf{G}_*\end{bmatrix},\quad \text{Joint Gaussian realization} \tag{18}$$

$$\mathbf{m}=\begin{bmatrix}\mathbf{m}_{\mathbf{G}}\\ \mathbf{m}_*\end{bmatrix},\quad \text{Joint mean vector}$$

$$\boldsymbol{\Sigma}=\begin{bmatrix}\boldsymbol{\Sigma}_{\mathbf{G}} & \boldsymbol{\Sigma}_{\mathbf{G}*}\\ \boldsymbol{\Sigma}_{\mathbf{G}*}^{\mathsf{T}} & \boldsymbol{\Sigma}_*\end{bmatrix},\quad \text{Joint covariance matrix}$$

in which,

$$\begin{aligned}[\boldsymbol{\Sigma}_{\mathbf{G}}]_{ij}&=\rho(x_i,x_j).\sigma_G^2,\\ [\boldsymbol{\Sigma}_*]_{ij}&=\rho(x_{*i},x_{*j}).\sigma_G^2,\\ [\boldsymbol{\Sigma}_{\mathbf{G}*}]_{ij}&=\rho(x_i,x_{*j}).\sigma_G^2.\end{aligned} \tag{19}$$

The learning procedure is about estimating the hyper parameters (unknowns) based on the observation and updating our knowledge using joint probability for drawing conclusion about the posterior predictive (i.e. mean vector and covariance matrix). In the next section a detailed parameter estimation is provided.

5.4 Parameter estimation and model performance

Parameter estimation is the objective of the learning procedure to build a model. Here, we denote the parameters that need to be estimated as hyper-parameter vector $\boldsymbol{\theta}=[\sigma_G,l,p]^{\mathsf{T}}$.[1] The three

[1]Depends on the correlation function, the parameter vector can be change. For instance, in the case of using squared exponential correlation function, the hyper-parameter vecto is $\boldsymbol{\theta}=[\sigma_G,l]^{\mathsf{T}}$

hyper-parameters are estimated based on the observation using the Bayes' rule. Thus, one can obtain the posterior PDF of the parameters given data $\mathcal{D} = \{(\boldsymbol{x}_i, g_i), \forall i = 1 : \mathtt{D}\}$ as:

$$\overbrace{f(\boldsymbol{\theta}|\mathcal{D})}^{\text{posterior}} = \frac{\overbrace{f(g_i|\boldsymbol{x}_i, \boldsymbol{\theta})}^{\text{likelihood}} . \overbrace{f(\boldsymbol{\theta})}^{\text{prior}}}{\underbrace{f(g_i)}_{\text{marginalization constant}}}, \tag{20}$$

Calculating the posterior is computationally demanding. Instead, one can maximize the log-likelihood.

$$\boldsymbol{\theta}^* = \underset{x}{\text{argmax}} \quad \overbrace{\ln f(g_i|\boldsymbol{x}_i, \boldsymbol{\theta})}^{\text{log-likelihood}}. \tag{21}$$

Such maximization for parameter estimation is known as Maximum Likelihood Estimation (MLE). One can use any gradient approach for MLE. The performance of the model is then be measured by it likelihood. As such, the better the model we have, the more likelihood we obtain. In another words, once the likelihood is maximized, it means that given the obtained parameters, we expect to have the observation with higher probability.

5.5 Meta model

The objective of the intelligent model is to remove the need for numerical simulation such as FEM and replace the model with data-driven one based on observation. Here, a detailed explanation of building such model, meta model or surrogate mode, is provided. Without loss of generality, in this section we focus on building one model based on one of the covariates (features), namely $peak \geq 60\%$. Fig. 33 shows the variation of such attibute along the pipe when it is exposed to fault. Our observations are the result of finite numerical analysis. As such, one may realize that in order to build a predictive model, we need finite observation. It is important to bare this fact in mind that as long as our observations are based on FEM model, observation noise, v_i, are equal to zero.

5.5.1 Gaussian Process for Machine Learning (GPML) tool

The Gaussian Process for Machine Learning (GPML) tool is a dedicated open source MATLAB® operated tool for Gaussian Process (GP). It was originally developed at the University of Cambridge machine learning group. The tool provides a neat software for GP including significant library of kernel functions. In the current research such tool is utilized and modified for civil engineering applications. There are primary key points that the reader needs to pay careful attention. Here we provide them[2] and the reader is referred to [30] and Gaussian Process Website for detailed explanation of using such a tool.

a) The observation noise is set to zero as they are induced from numerical analysis.

[2]The detailed MATLAB® codes are provided in the appendix.

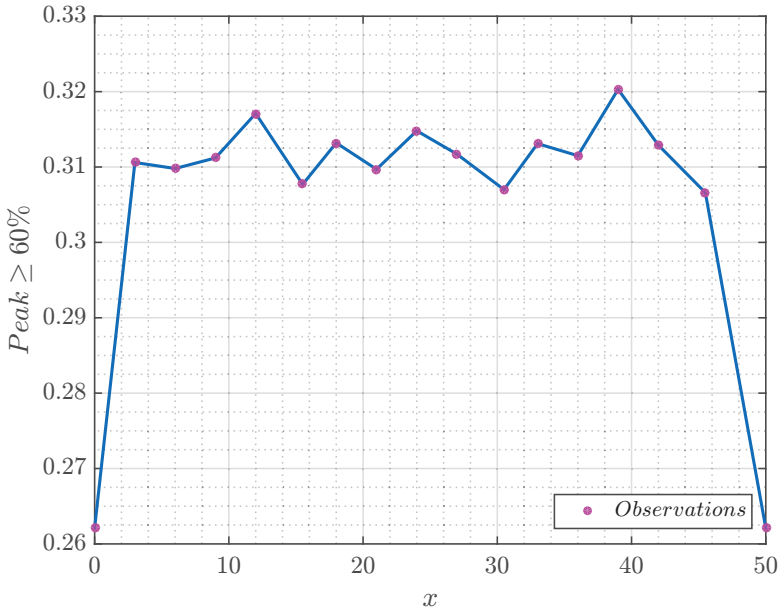

Figure 33: Feature 1.

b) The choice of correlation functions depends on the expected behavior of the covariates.

c) To prevent over-fitting, special consideration should be taken (see Section. 5.5.2).

5.5.2 Over-fitting

Over-fitting happens when the model works well on training data sets but fails to show good performance on other data sets that have not been seen by the model. To prevent such misleading results, one may use n-fold cross validation. In such an approach, the data set is divided into n folds. $n-1$ folds are used to train the model and the last fold is used for test. The whole procedure is repeated n times to make sure that each fold is used in the training and test sets. Finally, the results (estimated parameters) are averaged and reported to build the final model. It is common to use 5 or 10 folds. However, the best results are obtained when one increase the number of folds: the maximum will reach once the number of folds is equal to the number of instances. As such, the n-fold cross validation is called Leave-One-Out-Validation (LOOV). Nevertheless, once the size of data sets is significantly increased, LOOV is computationally demanding. In this study we use LOOV for building the model because the number of instances are small. It is worth noting again that working with small data sets (small sets of observations) is one of the primary advantages of the Gaussian Process.

5.5.3 Illustrative result

In this section we use one feature and investigate the meta model along with LOOV. Table. 4 shows the corresponding data to Fig. 33. For the sake of brevity, we name instances by their ID as shown

in the table. In addition, the first and last value is corresponding to the intact structure, so there is no difference in values for them. As previously mentioned, in LOOV procedure, we put aside one

Table 4: First feature ($peak \geq 60\%$) ID, values, and location of the defect.

ID	$peak \geq 60\%$	Location
1	0.2621	0
2	0.3106	3
3	0.3098	6
4	0.3112	9
5	0.3171	12
6	0.3077	15.5
7	0.3132	18
8	0.3097	21
9	0.3148	24
10	0.3117	27
11	0.3070	30.5
12	0.3131	33
13	0.3115	36
14	0.3203	39
15	0.3129	42
16	0.3066	45.5
17	0.2621	50

observation (instance) as we test, and train the model for the rest of the data. Afterwards, we use the test set to evaluate the performance of the model. Here, the metric for the performance is the likelihood of the prediction capacity of the model. To illustrate the procedure, we start from having, for instance, 5 observations (ID $\in$ [1,3,9,15,17]). Fig. 34 shows the corresponding result. As it is seen, the number of observation is not enough to form an applicable model. Although the prediction and uncertainty bandwidth are illustrated. In another words, one can draw results about damage detection and localization, but the uncertainty is high that we cannot be sure about our decision.

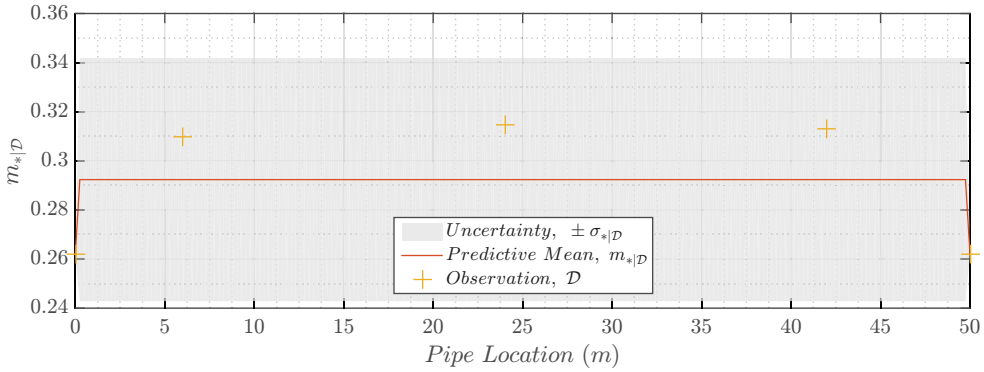

Figure 34: GP model for the pipeline with finite observations ID $\in$ [1,3,9,15,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0247, l \approx 0]^\mathsf{T}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) = 11.4$

Interestingly, we obtained $l =\approx 0$ for the squared exponential correlation function meaning that

the correlation function has more spikes and one can infer that the observation at each location cannot be correlated with the neighbor locations. This phenomenon is not due the fact that the selected feature is not appropriate although in some cases it is true. However, by increasing the number of observation, such uncertainty can be shrunk to the point that the engineer can draw conclusion to detect and localize the damage. Fig. 35 shows the effect of an increase in the observation. As it is seen, by not having observation in the middle part of the pipe the uncertainty of the model increases, while near the observation we face less uncertainty. In addition, Fig. 36 shows the predictive model corresponding to all observations except ID=14. It is seen that again having more observation results in better prediction. The sample MATLAB® code for this analysis is provided in the appendix.

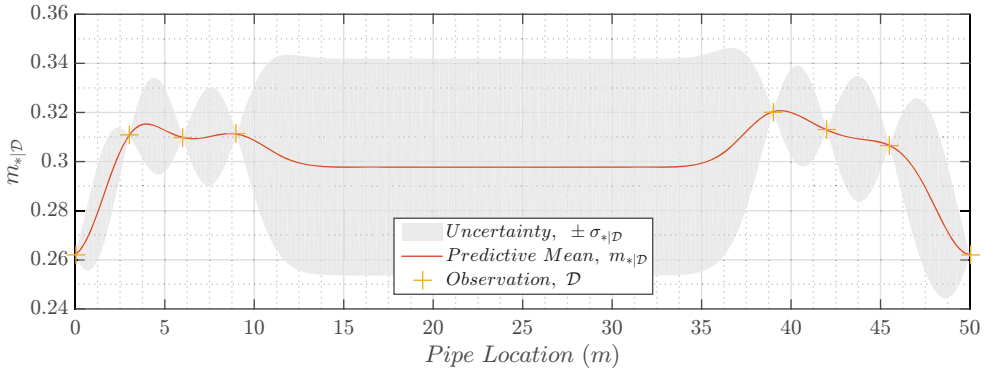

Figure 35: GP model for the pipeline with finite observations ID $\in$ [1,2,3,4,14,15,16,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0221, l = 1.7577]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) = 19.3$

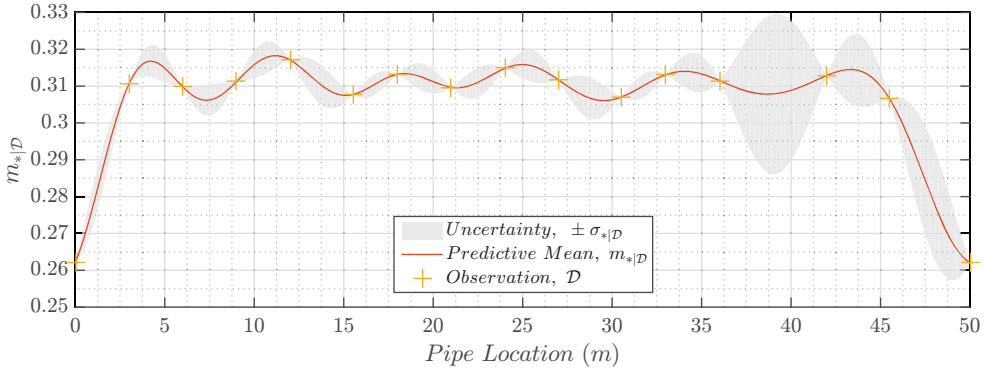

Figure 36: GP model for the pipeline with finite observations ID $\in$ [1,2,3,4,5,6,7,8,9,10,11,12,13,15,16,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0194, l = 2.7844]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) = 43.7$

To compare the performance of the model with respect to increasing the observations, one may look at the log-marginal-likelihood of the estimated parameter for selected feature (i.e. $\ln f(g_i|peak \geq 60\%, \sigma_G, l)$).One can see the obtained log-likelihood for the models as shown in the caption of each figure. As mentioned, the better the model we have, the higher log-likelihood we obtain. The

only question remains, is whether the model is over-fitted or not. As mentioned, to overcome this situation LOOV procedure is applied (see Fig. 36). Fig. 37 illustrates final results of the predictive model for selected feature.

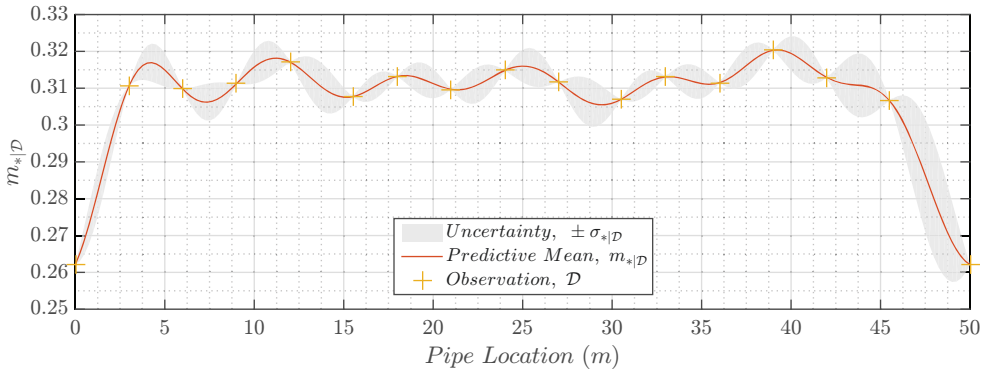

Figure 37: GP model for the pipeline with finite observations ID $\in$ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0188, l = 2.6545]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) = 46.7$

Chapter 6

Results, discussion, and comparison

In this chapter we extend the GPRM for the reset of the selected features. We examine the features for their capability for either damage detection and/or localization. Finally, we provide a comparison between GPRM and conventional regression.

6.1 Results and discussion

In this section we will discuss the other features of the Meta model.As we can see some features provide good results to find the location of fault (localization) or even if there is any fault or not (detection), and some others due to lack of number of observations may give false-positive results. Figs.38-43 illustrate the GP results. Unless otherwise expressed, in the following discussion the feature *peak* $\geq 60\%$ is excluded from the discussion.

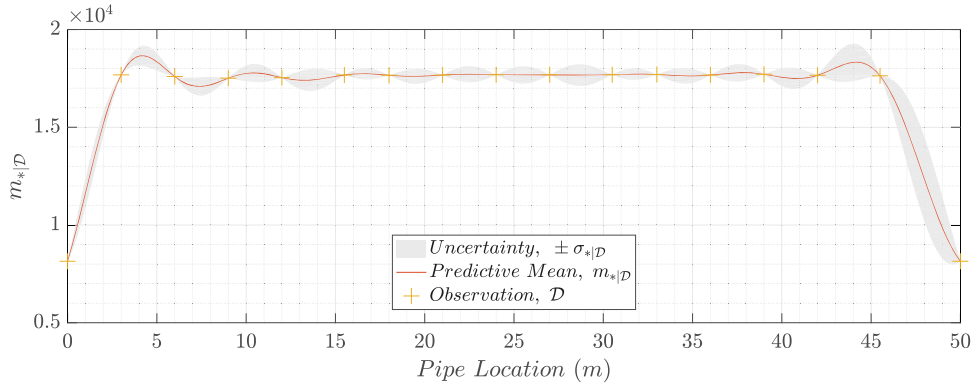

Figure 38: GP model for the pipeline with finite observations ID $\in$ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 4.1E3, l = 3.19]^{\mathsf{T}}$. $\ln f(g_i|CL, \sigma_G, l) = 46.7$

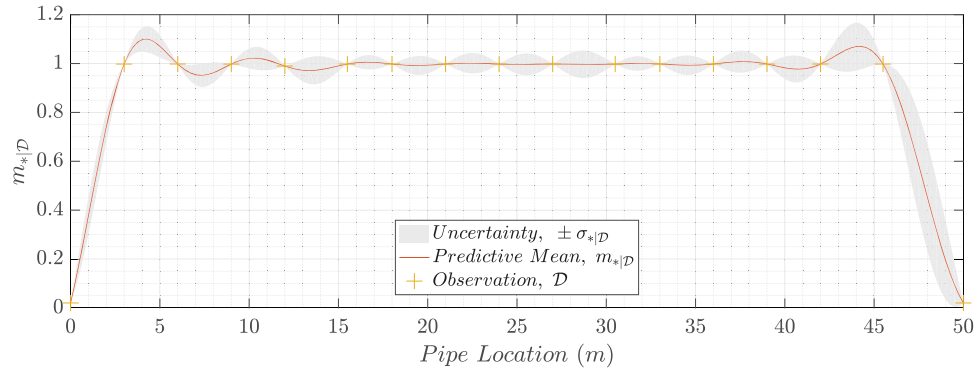

Figure 39: GP model for the pipeline with finite observations ID $\in$ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.42, l = 3.2]^{\mathsf{T}}$. $\ln f(g_i|COR, \sigma_G, l) = 48.45$

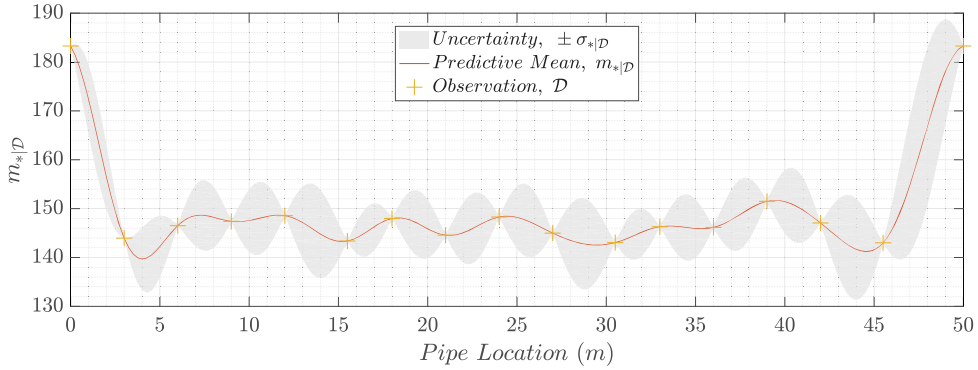

Figure 40: GP model for the pipeline with finite observations ID $\in$ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 13.03, l = 2.13]^{\mathsf{T}}$. $\ln f(g_i|KU, \sigma_G, l) = 45.6$

Figs. 38 - 42 show good *capability*[1] with respect to damage detection. The reason is that we can see a significant jump from intact to damaged pipeline. This jump can be as an identifier for damage detection. Nevertheless, one has to pay attention to the fact that there is no observation along such jump and, before having observation(s) within this range, it is recommended to use such graphs with enough cautions. In addition, Fig.38 and Fig.39 reveal less uncertainty between the two state of the pipe. As a result, these two features (i.e. curve length and correlation coefficient) are the best candidate for damage detection. It is worth noting that these two also show better damage detection predictabily compared with the feature *peak* $\geq 60\%$. On the other hand, non of the features corresponding to Figs. 38 - 42 show a significat capability towards damage localization. However, Fig. 40 (Kurtosis) shows relatively good capability for damage localization, but choosing features *peak* $\geq 60\%$ and *peak* $\geq 20\%$ are more promissing. Fig. 43 (*peak* $\geq 20\%$) illustrates powerful capability for damage localization, but not detection. As a general rule, one may use curve

[1]The term is used as whether the decision maker can easily make distinction between different states with less uncertainty, qualitatively. Such term has to be used with enough care so it will not lead to wrong decision. As an extension of the current study, one may use intelligent decision makings such as Markov Decision Process in order to quantify as well as automate such decision.

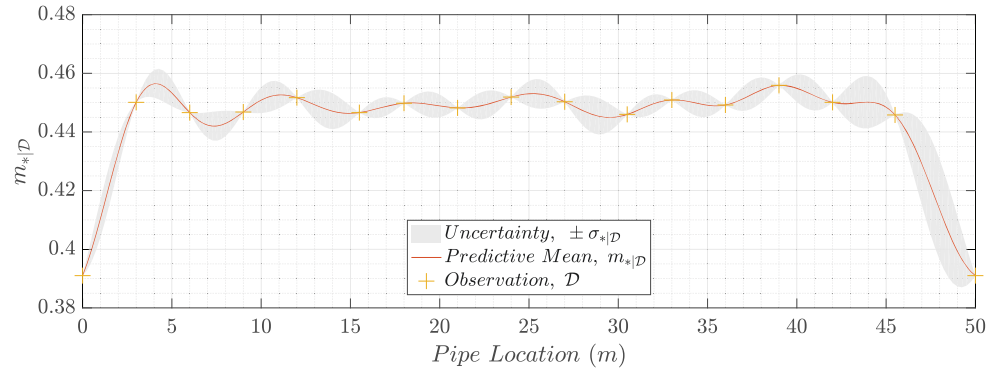

Figure 41: GP model for the pipeline with finite observations ID $\in$ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.227, l = 2.78]^{\mathsf{T}}$. $\ln f(g_i|L2, \sigma_G, l) = 43.2$

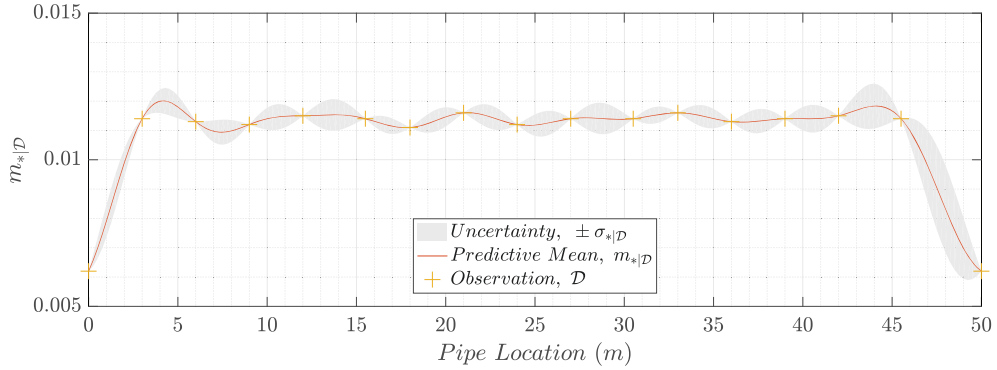

Figure 42: GP model for the pipeline with finite observations ID $\in$ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.002, l = 2.8]^{\mathsf{T}}$. $\ln f(g_i|MeanPeak, \sigma_G, l) = 45.8$

length and correlation coefficient (Fig.38 and Fig.39) for damage detection, while Fig. 43 and Fig. 37 ($peak \geq 20\%$ and $peak \geq 60\%$, respectively) provide predictive model for damage localization.

6.2 Comparison

In this section, we compare the obtained results from GP with those of conventional refression. To this end, we chose three different functions to create the predictive model, namely polynomial, Gaussian, and Sinusoidal[2]. In what follows a short description of the models are provided. The reader is referred to the available MATLAB® library for comprehensive explanations.

Exponential model In this study we utilized the exponential model of order 9, *poly9*, tha can be expressed as

$$g(x) = a_1 x^9 + a_2 x^8 + ... + a_9 x + a_{10} \tag{22}$$

[2]We follow the same notation for the function (i.e. $g(x)$) as of previous chapters for mapping the attribute(s) to the output (observation data)

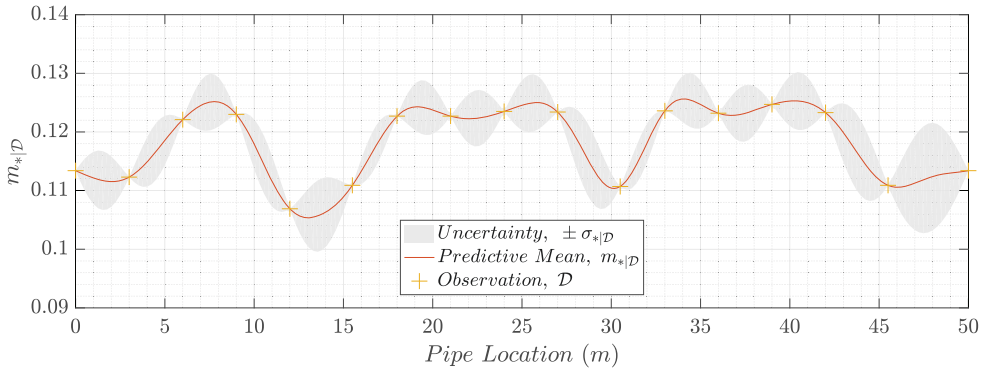

Figure 43: GP model for the pipeline with finite observations ID $\in$ [1,2,...,17]. The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 0.0063, l = 1.88]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 20\%, \sigma_G, l) = 46.9$

Gaussian model The next model that is used is the Gaussian type function, *gauss2*, which is written as

$$g(x) = a_1 \exp\left[-(\frac{x - b_1}{c_1})^2\right] + a_2 \exp\left[-(\frac{x - b_2}{c_2})^2\right] \tag{23}$$

It should be pointed out that the reader shouldn't confused this function with Gaussian process. Here, the shape of the function looks like Gaussian function.

Sinusoidal model The last function that is used is of the sinusoidal one , *sin3*, which is expressed as

$$g(x) = a_1 \sin(b_1 x + c_1) + ... + a_3 \sin(b_3 x + c_3) \tag{24}$$

In Eqs. 22 - 24, a_i, b_i, and c_i are the coefficients that are optimiezed by minimizing the corresponding cost functions (e.g. using gradient based algorithms). The reader is referred to the literature [15] for such optimizations. The results are shown in Fig. 44 for all three selected functions.

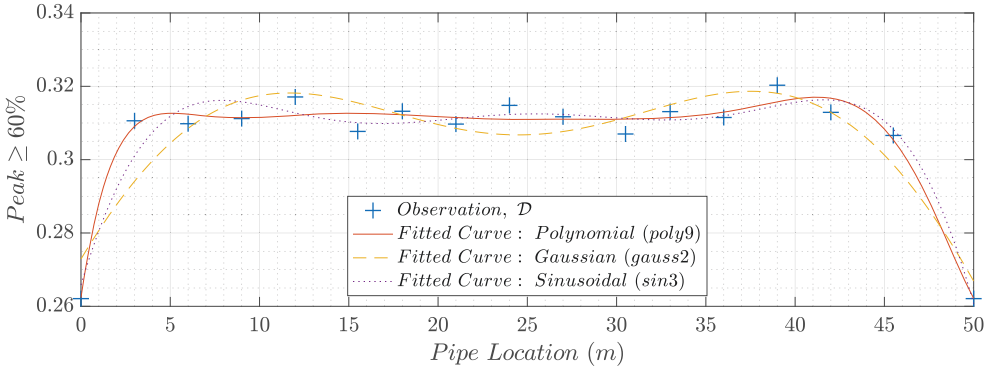

Figure 44: Comparison of the regression with three different models for $peak \geq 60\%$.

As it is seen, the choice of polynomial and sinusoidal models fails to predict the damage location and they just show significant change between the intact and damaged states of the pipeline. The Gaussian models illustrates better results compared with the other two models, but within some

areas (middle part of the pipe) it underestimate the results, while for the other parts it overestimate the system responses. Moreover, there is a question that *what types of functions should we choose?*. In fact, to answer this question the researcher (engineer) has to have a prior knowledge of the system response pattern(s). The choice of functions are almost infinite and unless otherwise no knowledge is available, one bare can choose the correct function. In addition, wrong choice of the function(s) may lead to poor results. Here, we chose the mentioned functions after many trials and errors. Another question rising here is *How does our belief change once we observe a phenomenon?*. The regression result curve has the deviation from the real observation. It is due to the fact that the regression analysis tries to minimize the total cost (error function) between all data points. This leads us to the concept of uncertainty. As it is expected in reality, once we observe a phenomenon, our uncertainty will reduce. In another words, our belief about the phenomenon will be stronger as we gain more knowledge about it. But, regression analysis does not address such issue, so it cannot differentiate between not observing or observing to the great extent—instead, it tries to minimize such deviations globally. For the sake of illustration, Fig.X shows the comparison between the three regression models and GPRM for $peak \geq 60\%$. It is worth noting that for the other features, once can draw the same conclusion.

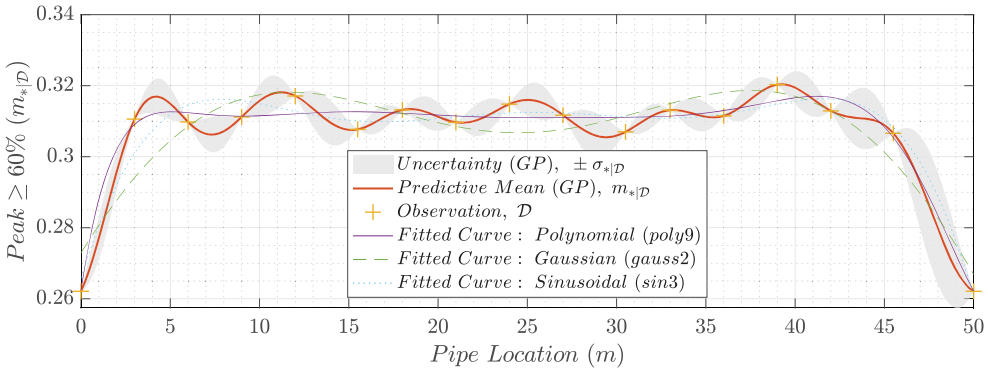

Figure 45: Comparison of the regression models and GPRM for $peak \geq 60\%$.

6.3 Decision making

Decision making is the last step in the previously introduced framework. Generally, one may use the curve length and correlation coefficient feature (and corresponding graph) for damage detection, while $peak \geq 20\%$ and $peak \geq 60\%$ can be utilized for damage localization. Moreover, one may use other time of intelligent models in order to automate the decision making, such as Markov Decision Process (MDP). Such intelligent framework is usually known as Reinforcement Learning (RL) and it is beyond the scope of the current study. Interested readers are referred to the literature for implementation of such a framework.

6.4 Special study

This section is dedicated re-examining the GPRM for a single 1.5m long wood beam. Here the wooden beam is tested and the GPRM is applied to the data to determine if the method works well. Although the contentious beam tested here is not a pipe, it can qualitatively represent the behaviour of a contiguously supported pipeline, and help demonstrate the applicability of GPRM in damage detection of such systems. The experiment is carried on to demonstrate the real-life application of such approach when we are dealing with obsrevational errors. In previous analysis, the meta model is used to build the intelligent model for prediction purposes. The difference here is the observation error, which previously was set to zero (See Chapter.5). Therefore, here we start with how to install the wood beam and its associated suspension system, the connector and the impact points are discussed. Finally, it is explained how the experiment is carried out. The wood beam is almost homogeneous, and the free conditions are considered on both sides of the wood. Accordingly, and to comply with the practical requirements of the test, a relatively soft spring is used to suspend the wood so that the added frequency to the system (rigid frequency) would be close to zero (see Chapter.3). The suspension system should be such that it does not affect the shape of the modes. According to the results obtained by the finite element method (Ansys software) as well as analytical methods (for the morphology of the mode shapes), the movement of each of the four shapes of the first mode is perpendicular to the wood axis and the suspension system is perpendicular to the direction of excitation.

The location of the supports should be on the mode shape nodes that is important. Given that the shape of the first mode is more important than other modes, the support connection location is selected in the first mode node. In this experiment, the wood beam is divided into three equal parts. An accelerometer is connected to the beam and the impact is applied with a hammer. This work is repeated three times at each point so that the results are averaged to eliminate the possible noises due to external effects. It should be pointed out that this procedure is simillar to the one presented previously in Chapters.2 and 3. In this test, the Larze technology[3] (accelerometer and data acquisition system) is used to transfer and store data. It should be noted that it is attempted to allow the beam to be perpendicular to the suspension system as much as possible. Failure to comply with this issue is the source of the error itself. Since the weight of the sensor used in this test is less than the total weight of the structure (about 5 g), the sensor weight does not affect the results.

After installing the beam and connecting the sensor and the hammer to the device, it is time to perform the initial processing of the data using the MATLAB software. Here the data vector range and the number of Fourier transform spectral lines (the number of frequencies investigated in the range) are determined. As indicated in the previous sections, the response time is received by the accelerometer and the device converts them using the Fourier transform to the frequency domain after digitization of the data. Here using the impact force data, we provide an extensive analysis to examine the selected feature (peak $\geq$60%). As it is seen, the current feature has disadvantages with

[3]The reader is referred to the Larze technology website at: *https://www.sensequake.com* for more information

respect to the uncertainty. Although the feature is sensitive it may not localize the damage.

In the first stage of modeling in Ansys software, first the cross-sectional area of the beam is determined and then the shell volume of the wood is applied along the beam. In case of applying the fault in the wood, a pit with known dimensions and in defined places is created. And in the next step the pipe properties are attributed to it. At the meshing stage, by choosing the square meshes with appropriate dimensions to create effective elements in the problem-solving process are selected and finally the analysis has reached the solution step. By performing modeling and analyzing incrementally, in order to reach the effective length of the analysis in a number of lengths, finally the acceptable length of one kilometer is achieved. Since the sampling frequency is twice the maximum working frequency, the sampling frequency is obtained by the results obtained from the analysis. Due to the lack of computational facilities for the analysis of wood, it is decided to reduce the length of the pipe with regard to the available facilities. Accordingly, the results of the experiments in this thesis are based on 1.5m wood. Accordingly, it is concluded that it is possible to detect the fault location throughout the wood. Due to the simulation limitation and lack of hardware facilities, 3 faults at intervals of about 0.5meters from each other are created and results of each acceleration are obtained separately at the end of the wood. All the applied faults are along the length of the wood. In order to understand the effect of faults in the wood, it is first necessary to prepare accelerated data. Then the obtained data are converted from the time domain to the frequency domain.

One of the common methods for identifying the systems properties (e.g. stiffness or modal properties) of a structure is to test them under static or dynamic loadings. To identify the dynamic systems properties (e.g. stiffness or modal properties) of a structure, and elasto dynamic system properties, the modal analysis is performed on them.The purpose of this experiment is to obtain the natural frequencies of the pipe and the shape of the corresponding modes. In this experiment, first the natural frequencies and the shape of the pipe modes are obtained using a finite element software, Ansys. Using finite element analysis, the location of the nodes and pipe installation are determined. The results of pipe modeling and modal analysis of the pipe in Ansys software are presented below.

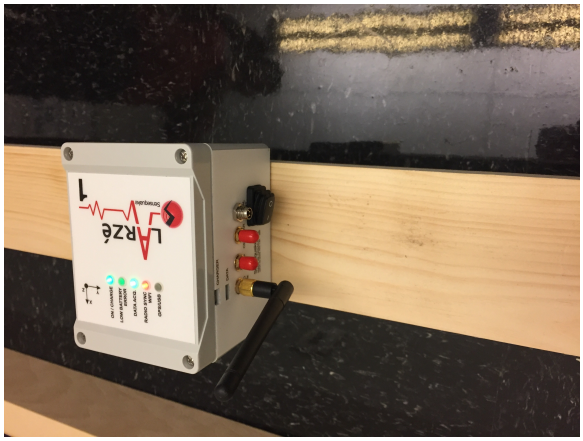

Figure 46: Larze technology (accelerometer) located on wood

Figure 47: Larze technology (accelerometer) located on wood

Figure 48: Fault on the wood beam

Fig. 50 shows the results obtained from GPRM. As it is seen the observation error is around 10%. However, the results shows that we need more observation for fault detection and localization. Figure 50 indicates that there is high uncertainty between the observation points and significant fluctuation at the end of the beam (60 to 100% length). Therefore, it is important to fine-tune the model by adding more observation points and validate it with further experimental study.

Figure 49: Fault on the wood beam

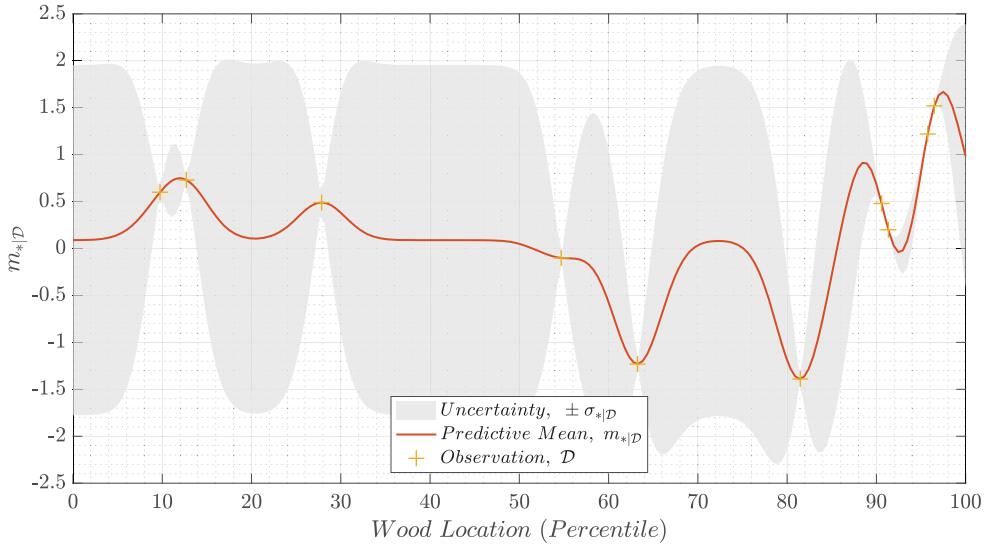

Figure 50: GP model for the wood The optimized parameter are $\boldsymbol{\theta} = [\sigma_G = 2.6, l = 5, \sigma_V = 0.1]^{\mathsf{T}}$. $\ln f(g_i|peak \geq 60\%, \sigma_G, l) = 10.5$

6.5 Summary

In this chapter we provide an extensive analysis to examine the selected features as well as perform a comparison between the current framework with conventional method. As it is found, the current framework has advantages with respect to considering the uncertainty and facilitating the choice of function for performing the regression. It is found out that although features are sensitive to the introduced damges, some of them can be utilized for damage localization.

Chapter 7

Conclusion and future work

7.1 Conclusion

In this study an intelligent approach was introduced to tackle the damage detection and localization along a 50m long pressurized pipeline. We proposed Gaussian Process Regression Method (GPRM) to establish dependency between various observations and create a probabilistic learnable model. To this end, from an available experiment, a meta-model was created in which the pipeline suffers from different defects (modeled as a small hole). Afterwards, sensed data were used to build a damage-sensitive library for the intelligent agent. Finally, the agent learned the corresponding parameter to create a predictive model. Three major results were obtained from this research: (i) GPRM has a great potential for monitoring of pipelines. The main reason is that it uses the observations to establish the dependency between them. It is evident that lack of observation at different locations increases the uncertainty around those locations. GPRM can handle such uncertainties by using correlation functions as in the vicinity of the observation the model shows less uncertainty and as the distance from the observed location increases, the uncertainty will rise. The second reason, is due to the nature of Bayes' rule to update the model in light of new observation. Pipelines are monitored periodically and during each monitoring process, new observations are added to the model and the agent can subsequently update its knowledge (posterior probability). (ii) The choice of correlation function plays an important role in the predictive model. In this study after various trials, squared exponential correlation function was found the most suitable one for the problem in hand, and (iii) Selecting proper feature(s) is crucial for both detection and identification. As it is seen all selected feature shows a huge discrepancy between the intact and damaged pipeline. However, just two of them ($peak \geq 20\%$ and 60%) were useful for damage localization. It is important to bear in mind that the selected features might be unique for the proposed problem and by changing the condition of the pipeline with respect to geometry, material, types of damage, and monitoring system, another set of features show more sensitivity to the damages.

Besides, by comparison with conventional regression analysis, GPRM reveals better results. First of all, the model is probabilistic so both model and observation uncertainties can be modeled. Secondly, the model is updatable by introducing new observations without introducing global effects

on the model. In contrast, in conventional regression, introducing new observation(s) may result in a totally new model. Finally, the focus of the GPRM is to establish the dependency along the pipeline by utilizing the correlation function, which is one function. However, the focus in regression is to minimize the error introduced by the selected functions. Such different behavior leads to the fundamental question about proper selection of choosing the function for the regression because wrong choice of functions leads to poor results. Long-term time series prediction is very useful in many application domains, yet it is a challenging task to perform. As time series prediction becomes our problem, we might want to use the Gussian Regression, which is one of the most prominent time series prediction models. we turn our attention to the Gaussian process regression, a non-parametric probabilistic regression technique. Due to its non-parametric property, a Gaussian process regression possesses a more flexible modelling capability than another model. With this flexibility, we hypothesise that the GP regression can be a more potent solution for long-term time series forecasting problem.

7.2 Future work and recommendation

This study shows the effectiveness of the GPRM for health monitoring of the pipelines. In this study we initiated a novel approach to address different aspects of monitoring (here is damage detection and localization). We can extend this work in different ways. Here we provide some possible yet promising extension of this study in three directions, namely "model enhancement", "detection and localization enhancement" and "Intelligent decision making".

- *Model enhancement*: The extension of the model can be categorized in two manners, namely physical and intelligence. From the physical perspective once can perform such analysis on the full-scale experiments (field-experiment) or enhance the numerical model with respect to different non-linearities and other types of damages such as additional mass. As such it is recommended that first the available numerical model is developed further to examine the effect of non-linearities, and then field experiment can be carried out to enrich the model with more realistic observations including the observational uncertainties. On the other hand, from intelligence vantage point, it is recommended to perform the Gaussian process in higher dimension. This study is limited to one-to-one mapping of features to system responses. However, one might interest in mapping more attributes to system responses. For instance, building an intelligent agent that can be learned from both $peak \geq 20\%$ and 60% at the same time.

- *Detection and localization enhancement*: Detection and localization were performed with limited number of features. Yet, there are numerous ways to create more damage-sensitive features. For example, one can use transformations such as Melin, Hilbert, Wavelet, and so on and then extract statistical features from transformed space. Having a vast number of features helps the researchers to understand the effect of other types of damages. Also, external effects may affect some of these damages and as such, having access to the bigger library of damages can help us to enrich the agent furthermore.

- *Intelligent Decision making*: In this study, after building the intelligent agent, the decision-making process is left to the engineers' judgments. However, such decision-making can be automated and/or intelligent by using other advanced method such as Markov Decision Process in the context of Reinforcement Learning (RL). Having another agent to make decisions can be helpful in large-scale applications where thousands of miles of pipelined should be monitored frequently.

In summary, utilizing GPRM in the field of pipeline health monitoring is promising in order to create intelligent agents to tackle various aspects of health monitoring including damage detection, localization, and decision makings. This study opened new areas of studies and improvements towards monitoring of large-scale structures and we hope that in near future, we would be able to see significant progresses in intelligent monitoring.

References

[1] M. A. Abdo. *Structural health monitoring: history, applications, and future*. New York:Open Science Publishers., 2014.

[2] R. J. Barthorpe. On model- and data-based approaches to structural health monitoring. 2011.

[3] F. Chang C. Boller and Y. Fujino. *Encyclopedia of Structural Health Monitoring*. New York: Open Science Publishers., 2009.

[4] M. Frigo C. S. Burrus, I. Selesnick M. Pueschel and S. G. Johnson. *Fast Fourier Transforms*. Connexions online book, Rice University, TX, USA, 2008.

[5] D. L. Margolis D. C. Karnopp and R. C. Rosenberg. *System Dynamics: Modeling, Simulation, and Control of Mechatronic Systems*. John Wiley and Sons Ltd, 2012.

[6] L. Ding and S. Liao. An approximate approach to automatic kernel selection. *IEEE Transactions on Cybernetics*, 47(3):554–565, 2017.

[7] D. J. Ewins. *Modal Testing: Theory, Practice, and Application*. Wiley, USA, 2009.

[8] C. R. Farrar and K. Worden. An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A*, 365:303–315, 2007.

[9] C. R. Farrar and K. Worden. *Structural health monitoring: a machine learning perspective*. New York: John Wiley and Sons, Inc., 2013.

[10] Intelligent Sensing for Innovative Structures (ISIS), Monitoring Structural Assessment, and Control (SAMCO). Educational module 5: An introduction to structural health monitoring, 2006.

[11] M. I. Friswell. Damage identification using inverse methods. *Philosophical Transactions of the Royal Society A*, 365:393–410, 2007.

[12] Kenya Fukushima, Reiko Maeshima, Akira Kinoshita, Hitoshi Shiraishi, and Ichiro Koshijima. Gas pipeline leak detection system using the online simulation method. *Computers and Chemical Engineering*, 24(2):453 – 456, 2000.

[13] Jinzhe Gong, Aaron C. Zecchin, Angus R. Simpson, and Martin F. Lambert. Frequency response diagram for pipeline leak detection: Comparing the odd and even harmonics. *Journal of Water Resources Planning and Management*, 140(1):65–74, 2014.

[14] L. H. C. He and B. Wu. Application of homodyne demodulation system in fiber optic sensors using phase generated carrier based on labview in pipeline leakage detection. 2006.

[15] E. Frank I. H. Witten and M. A. Hall. *Data mining: practical machine learning tools and techniques*. Massachusetts: Morgan Kaufmann Publishers, 2011.

[16] Pedro J. Lee, Martin F. Lambert, Angus R. Simpson, John P. Vítkovský, and James Liggett. Experimental verification of the frequency response method for pipeline leak detection. *Journal of Hydraulic Research*, 44(5):693–707, 2006.

[17] S. Z. M. Liu and D. Zhou. Fast leak detection and location of gas pipelines based on an adaptive particle filter. 42(10):54, 2008.

[18] G. Atia M. Malekzadeh and F. N. Catbas. Performance-based structural heath monitoring through an innovative hybrid data interpretation framework. *Journal of Civil Structure Health Monitoring*, 5:287–305, 2015.

[19] W. Peterson M. Mohitpour, T. Van Hardeveld and J. Szabo. Pipeline operation and maintenance: a practical approach, second edition. *ASME Press, New York*, 365, 2010.

[20] D. G. Manolakis and V. K. Ingle. *Applied digital signal processing*. Cambridge University Press, UK, 2011.

[21] S. Mayo, N. Money, C. Capus, and Y. Pailhas. Method and apparatus for acoustic assessment of fluid conduits, January 3 2014. WO Patent App. PCT/US2013/048,785.

[22] T. M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997.

[23] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Massachusetts Institute of Technology, 2012.

[24] R. M. Neal. *Bayesian Learning for Neural Networks. Lecture Notes in Statistics 118*. Springer, New York, 1996.

[25] Association of Oil Pipeline (AOPL). About pipelines., 2018.

[26] H. Ogai and B. Bhattacharya. *Pipe Inspection Robots for Structural Health and Condition Monitoring*. Springer (India) Private Ltd., 2018.

[27] S. Saha P. Tan P. M. Comar, L. Liu and A. Nucci. *Combining supervised and unsupervised learning for zero-day malware detection*. Proceeding of IEEE, 2013.

[28] J. K. Nisbett R. G. Budynas and J. E. Shigley. *Shigley's mechanical engineering design*. McGraw-Hill Education, New York, NY, USA, 2015.

[29] P. E. Hart R. O. Duda and D. G. Stork. *Pattern classification.* New York: John Wiley and Sons, Inc., 2001.

[30] C. E. Rasmussen and K. I. Williams. *Gaussian Process for Machine Learning.* MIT Press, Massachusetts Institute of Technology, 2006.

[31] G. Shmueli. To explain or to predict? *Statistical Science*, 25(3):289 – 310, 2010.

[32] S. W. Smith. *The Scientist and Engineer's Guide to Digital Signal Processing.* California Technical Publishing, San Diego, CA, USA, 1999.

[33] E. M. Stein and G. L. Weiss. *Introduction to Fourier Analysis on Euclidean Spaces.* Princeton University Press, USA, 1971.

[34] N. G. Thompson. E-gas and liquid transmission pipelines. *Technologies Laboratories, Inc.,Dublin, Ohio*, 2004.

[35] Shuli Wang, Mingming Lian, Shidong Zhou, Jianguo Feng, and Yongchao Rao. *Research and Outlook on Leak Detection Technology of Natural Gas Pipelines.*

[36] Xiao-Jian Wang, Martin F. Lambert, Angus R. Simpson, James A. Liggett, and John P. Vítkovský. Leak detection in pipelines using the damping of fluid transients. *Journal of Hydraulic Engineering*, 128(7):697–711, 2002.

[37] Larry Wasserman. Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, 44(1):92 – 107, 2000.

[38] S. Jin Y. Zhou and Z. Qu. Study on the distributed optical fiber sensing technology for pipeline leakage protection. *Proc.SPIE*, 6344:6344 – 6344 – 6, 2006.

[39] J. H. Garrett I. J. Oppenheim J. B. Haley J. Shi Ying, Y. and Y. Jin. Toward dta-driven structural health monitoring: application of machine learning and signal processing to damage detection. *Journal of Computing in Civil Engineering*, 27:667 – 680, 2013.

[40] J. Zhang. *Designing a cost-effective and reliable pipeline leak-detection system.* Springer, 1997.

[41] Y. Zhang and Y. Yang. Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, 187(1):95 – 112, 2015.

Appendices

Appendix A

Simulated time history data for the pipe

This chapter is dedicated to showing the figures of frequency due to no fault and fault at each 3 meter.

A.1 Section 1

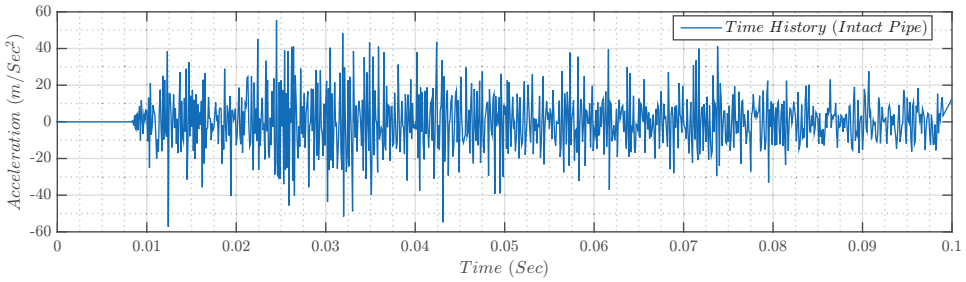

Figure 51: Time series of the pipe line with no fault (intact pipe).

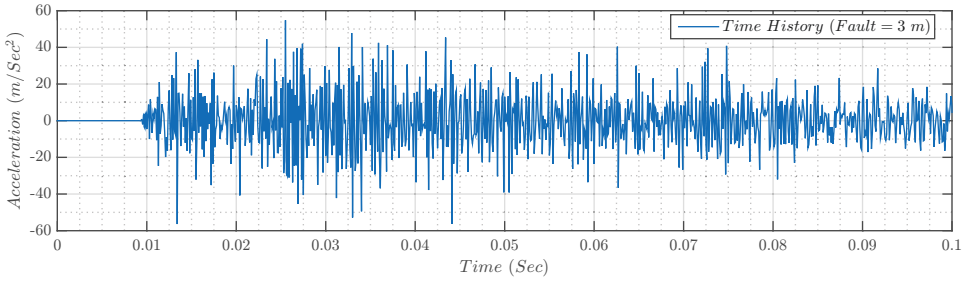

Figure 52: Time series of the pipe line with a fault located located at 3 m.

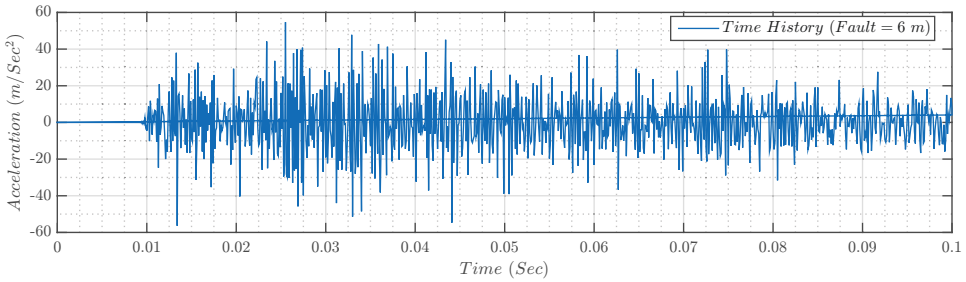

Figure 53: Time series of the pipe line with a fault located located at 6 m.

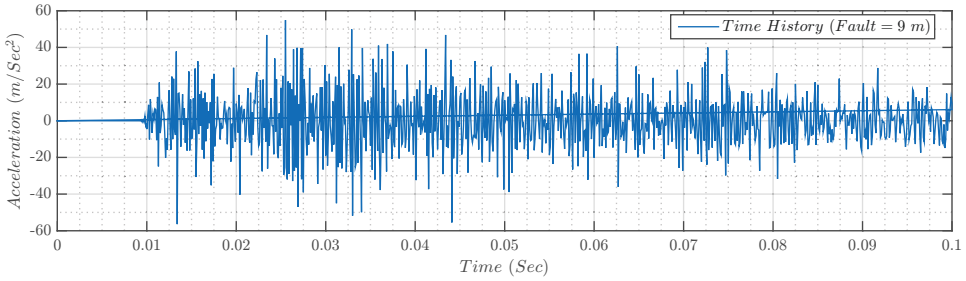

Figure 54: Time series of the pipe line with a fault located located at 9 m.

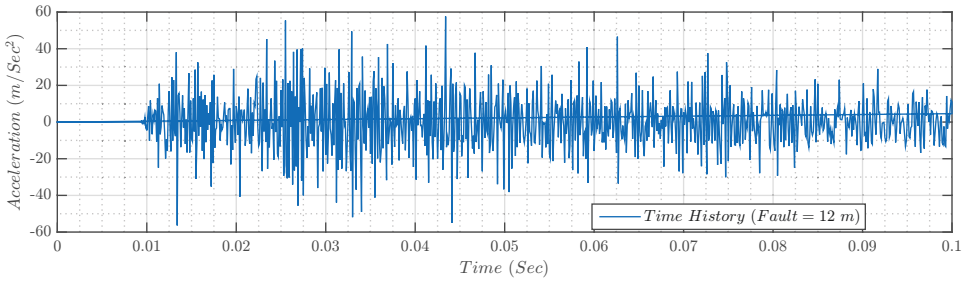

Figure 55: Time series of the pipe line with a fault located located at 12 m.

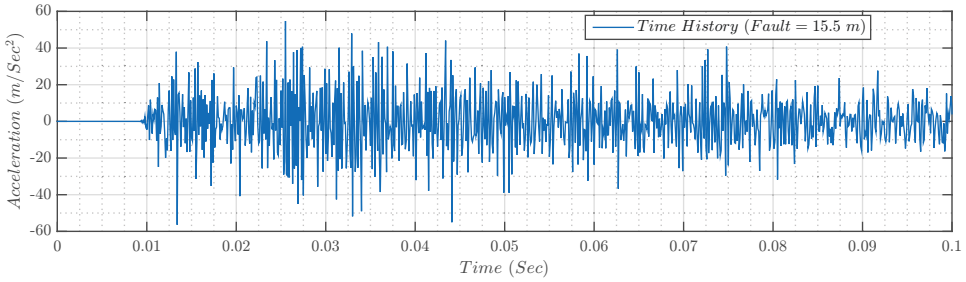

Figure 56: Time series of the pipe line with a fault located located at 15.5 m.

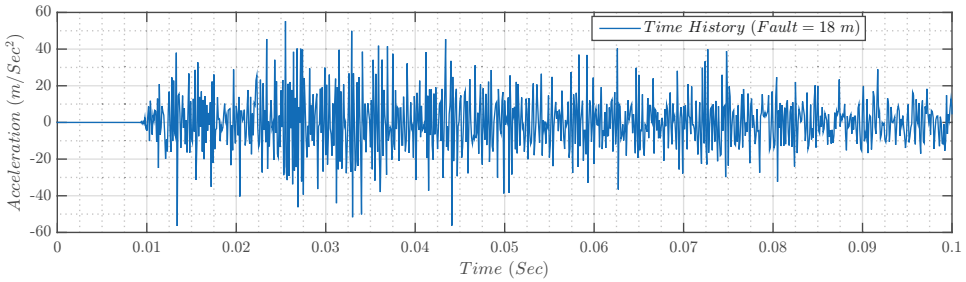

Figure 57: Time series of the pipe line with a fault located located at 18 m.

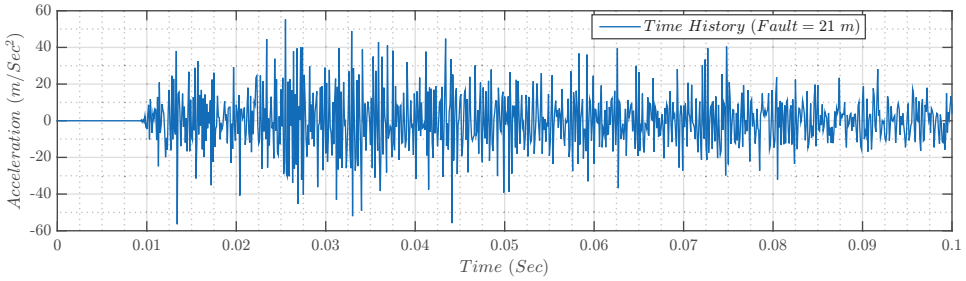

Figure 58: Time series of the pipe line with a fault located located at 21 m.

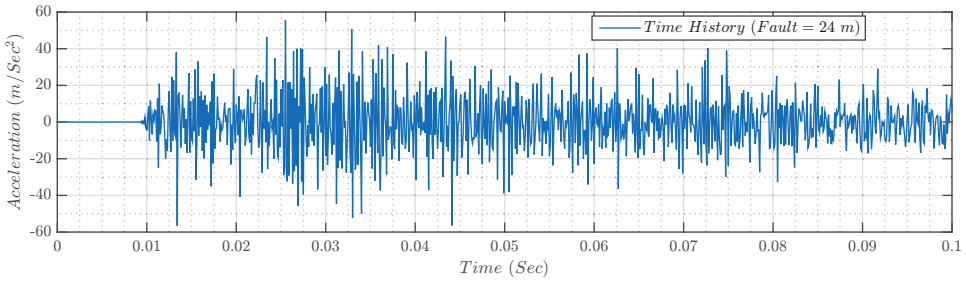

Figure 59: Time series of the pipe line with a fault located located at 24 m.

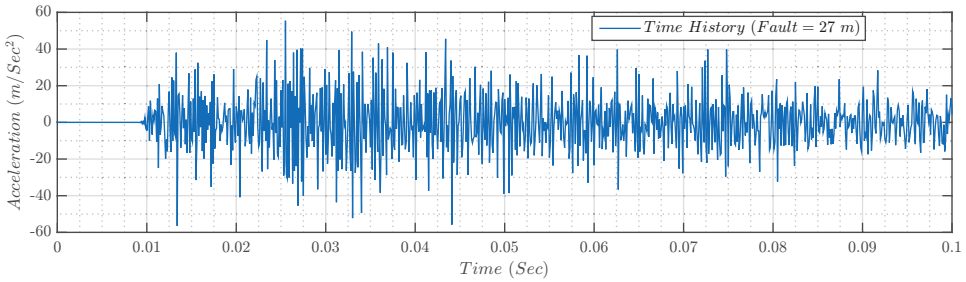

Figure 60: Time series of the pipe line with a fault located located at 27 m.

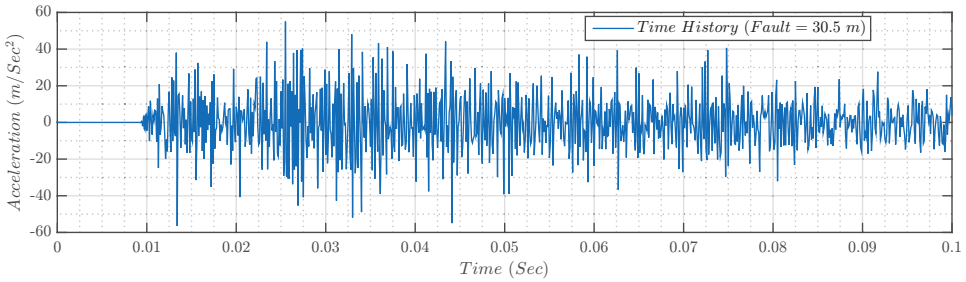

Figure 61: Time series of the pipe line with a fault located located at 30.5 m.

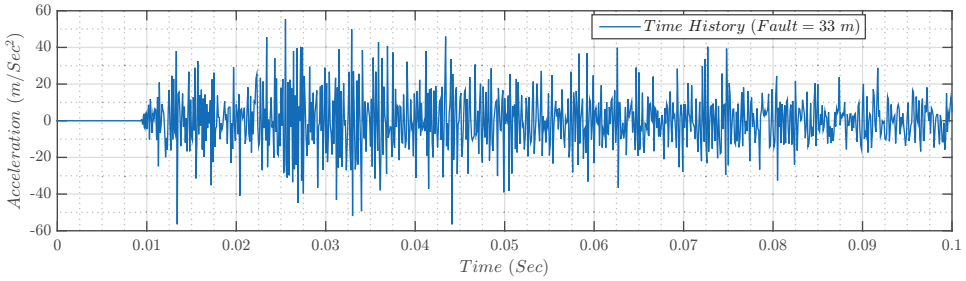

Figure 62: Time series of the pipe line with a fault located located at 33 m.

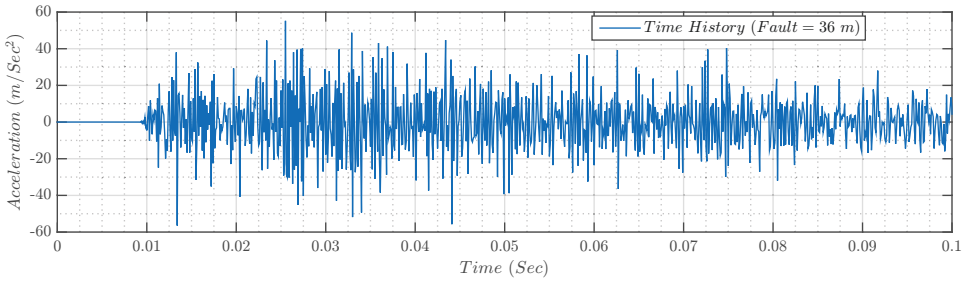

Figure 63: Time series of the pipe line with a fault located located at 36 m.

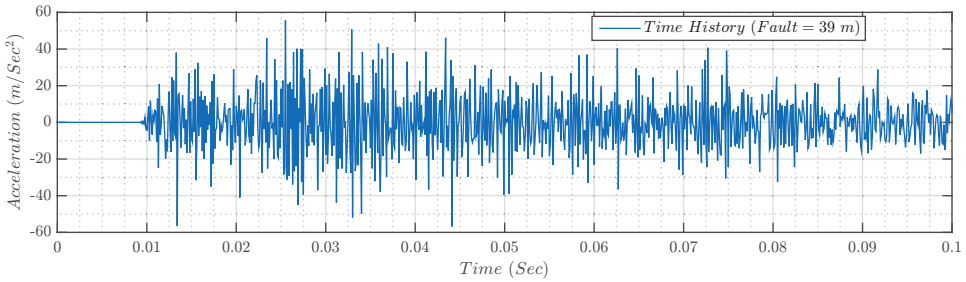

Figure 64: Time series of the pipe line with a fault located located at 39 m.

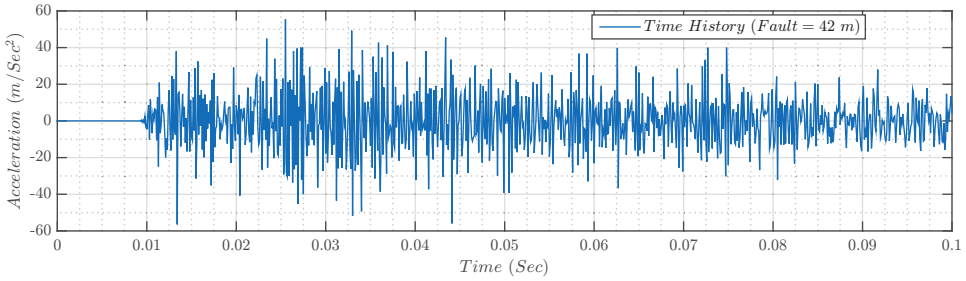

Figure 65: Time series of the pipe line with a fault located located at 42 m.

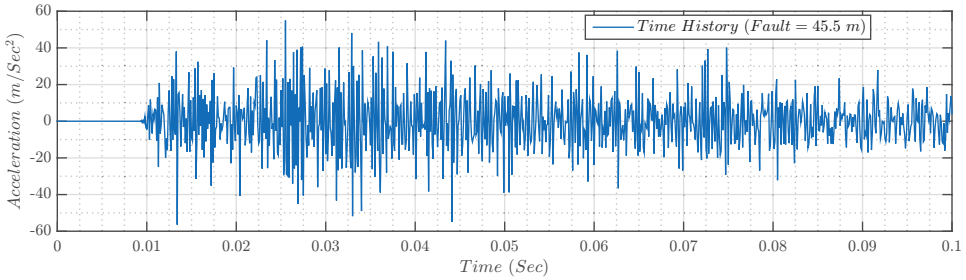

Figure 66: Time series of the pipe line with a fault located located at 45.5 m.

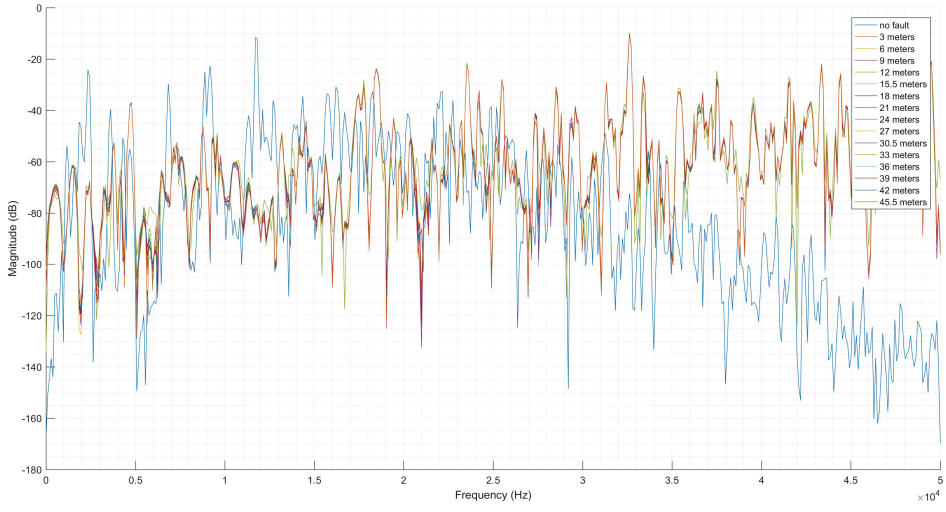

Figure 67: All Frequencies .

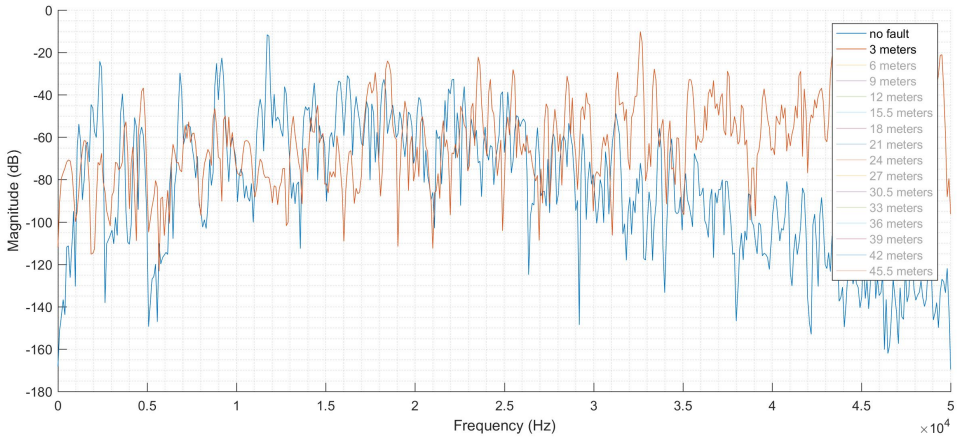

Figure 68: Frequency (No Fault vs. Fault at point 3m).

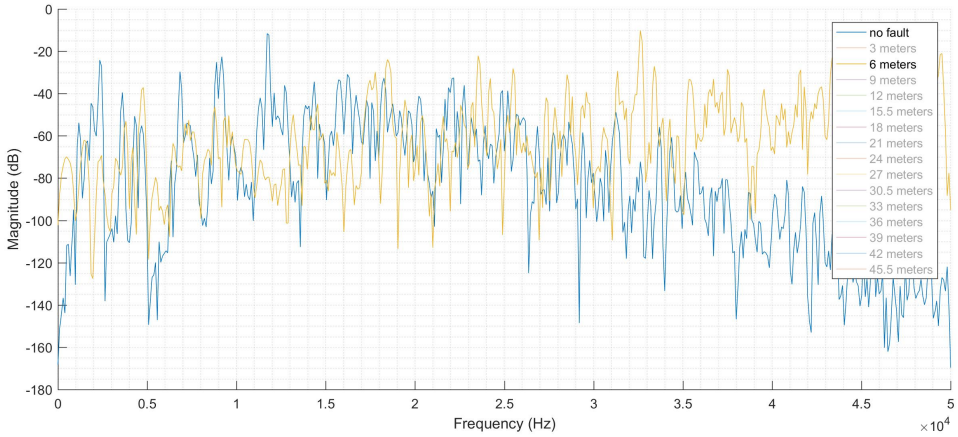

Figure 69: Frequency (No Fault vs. Fault at point 6m).

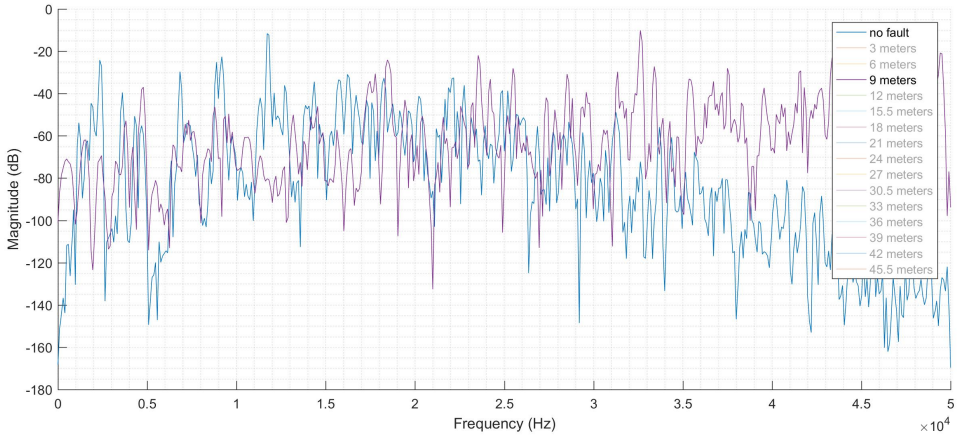

Figure 70: Frequency (No Fault vs. Fault at point 9m).

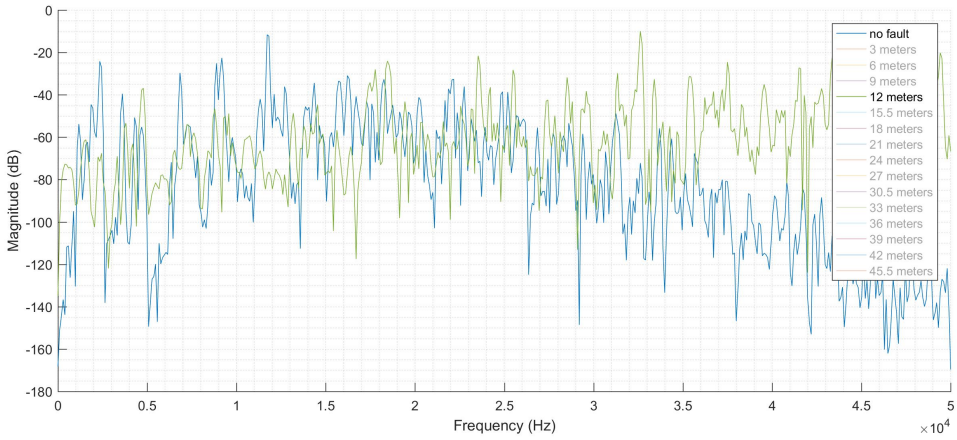

Figure 71: Frequency (No Fault vs. Fault at point12m).

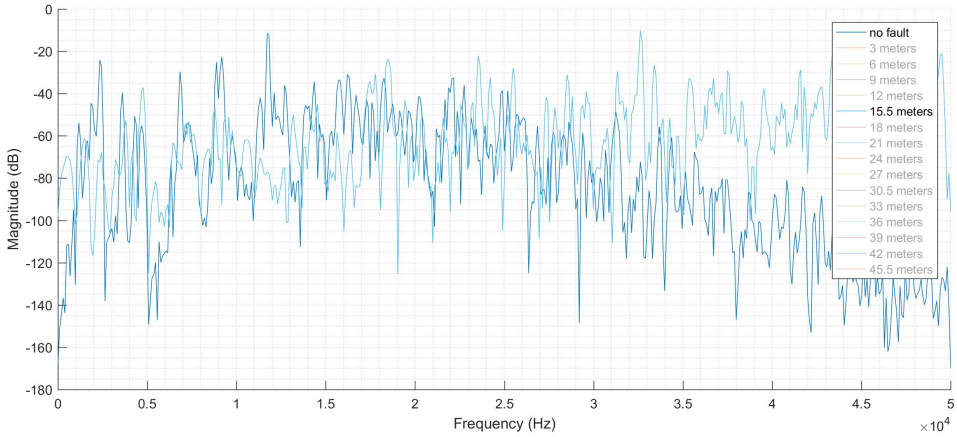

Figure 72: Frequency (No Fault vs. Fault at point 15.5m).

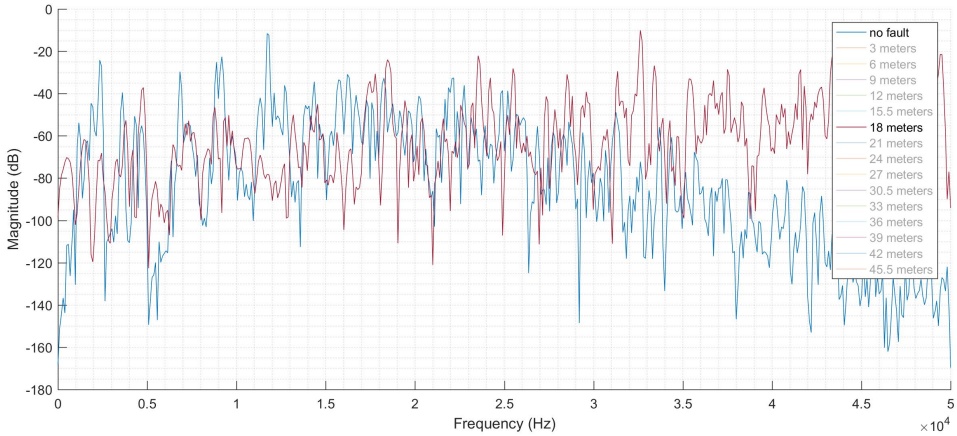

Figure 73: Frequency (No Fault vs. Fault at point 18m).

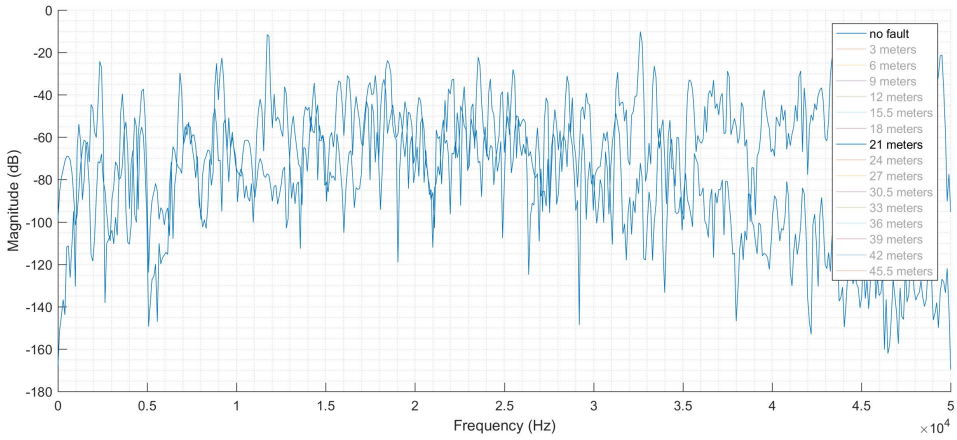

Figure 74: Frequency (No Fault vs. Fault at point 21m).

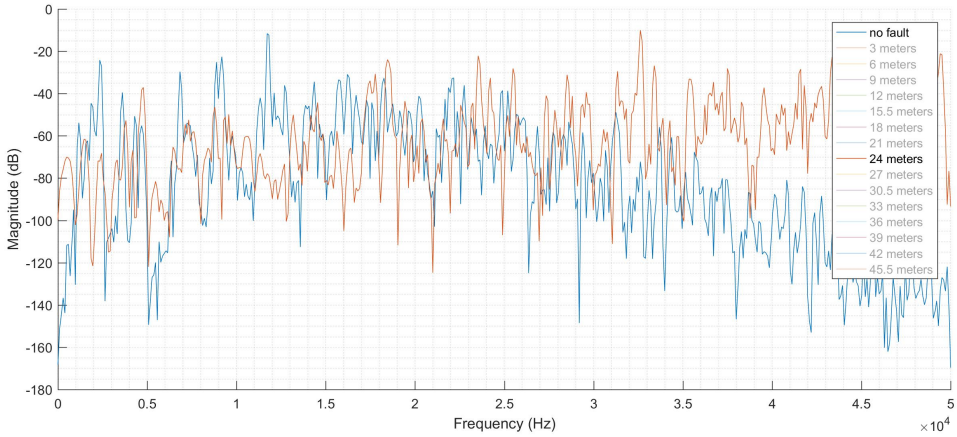

Figure 75: Frequency (No Fault vs. Fault at point 24m).

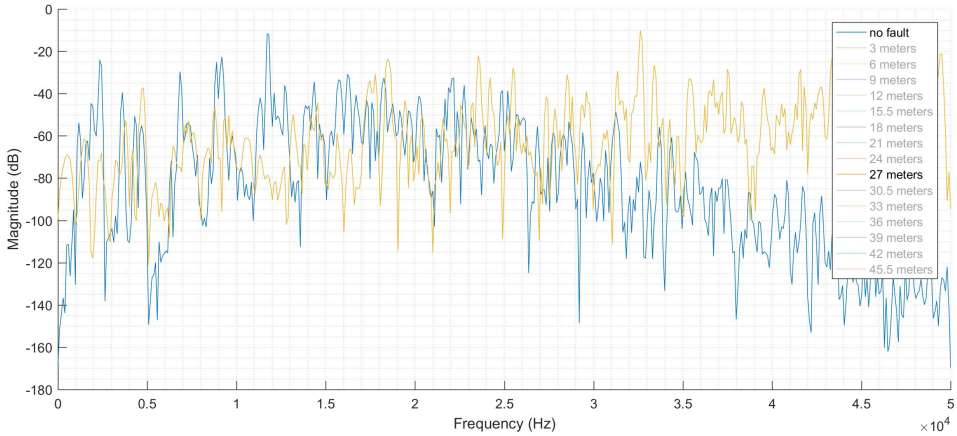

Figure 76: Frequency (No Fault vs. Fault at point 27m).

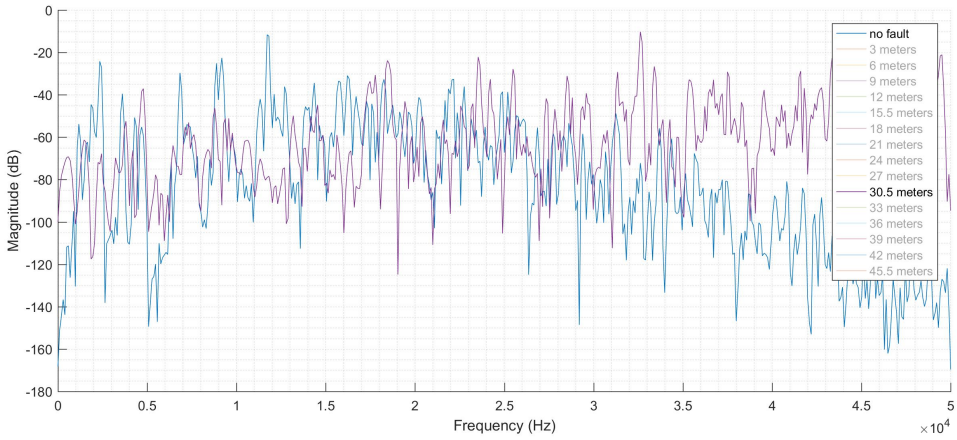

Figure 77: Frequency (No Fault vs. Fault at point 30.5m)

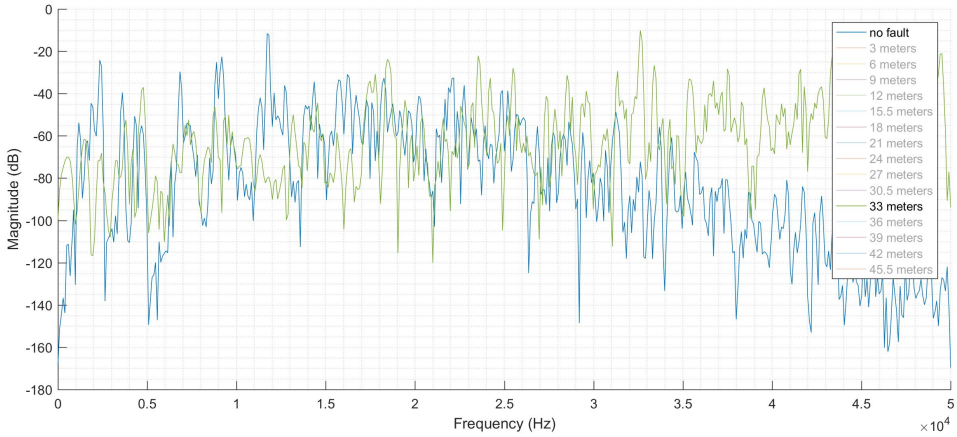

Figure 78: Frequency (No Fault vs. Fault at point 33m)

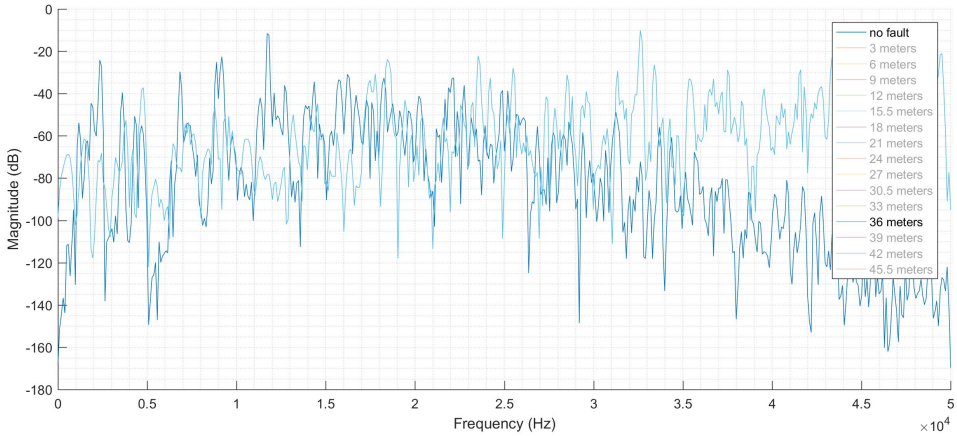

Figure 79: Frequency (No Fault vs. Fault at point 36m).

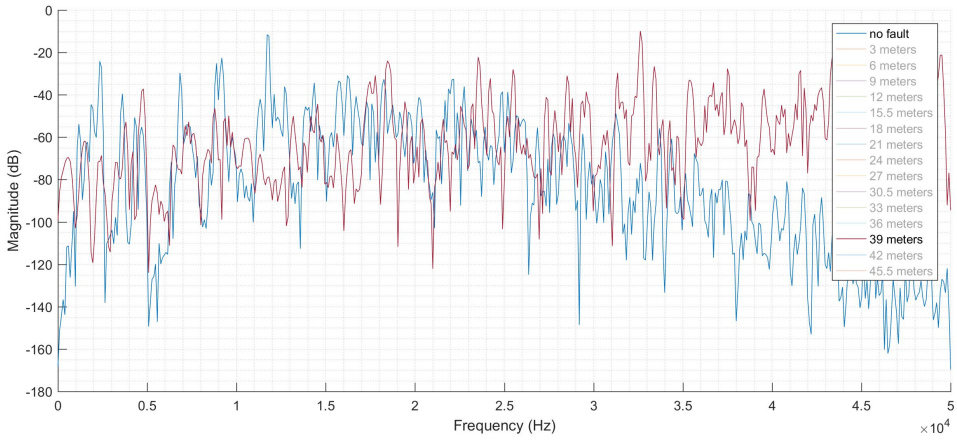

Figure 80: Frequency (No Fault vs. Fault at point 39m).

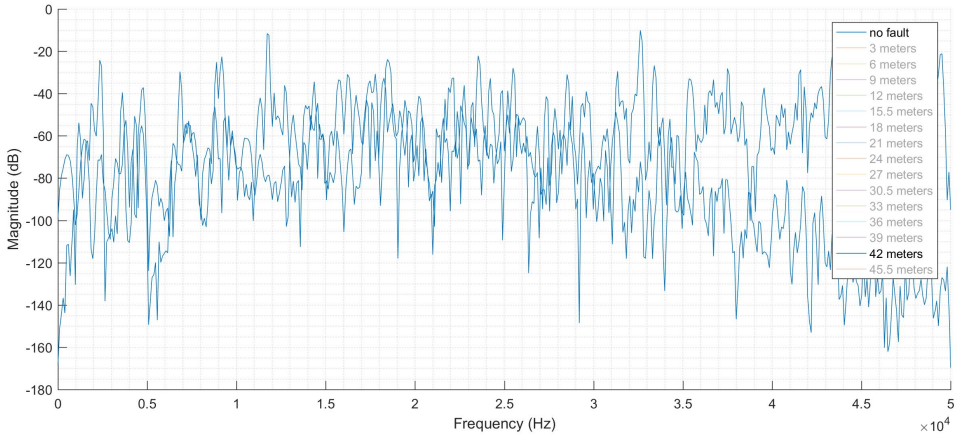

Figure 81: Frequency (No Fault vs. Fault at point 42m).

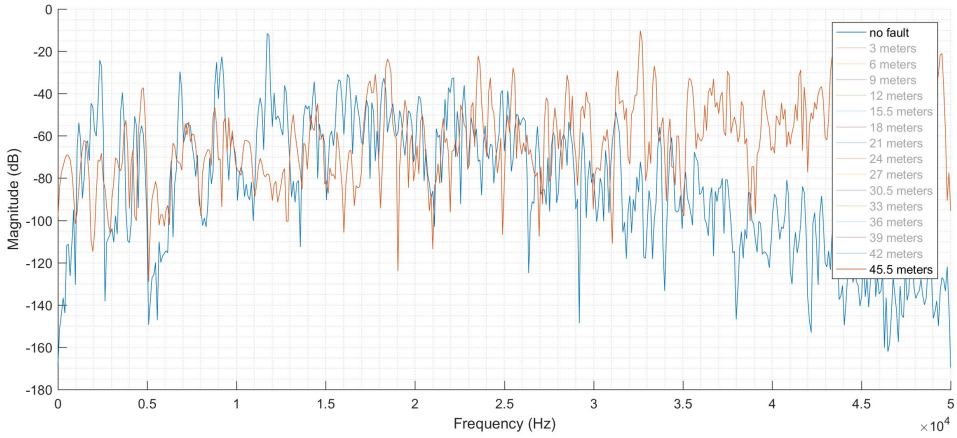

Figure 82: Frequency (No Fault vs. Fault at point 45.5m).

Appendix B

MATLAB® codes

Additional discussions and conclusions are provided at this chapter. Moreover, possible future works are expressed. As a special case, the use of the human brain as the pattern recognition tool is examined here.

B.1 Chapter 3

MATLAB®

```
close all; clear; clc;
x_obs = 0:0.001:1;
L = 1;
k_n = 1/L * [4.7300, 7.8532, 10.9956, 14.1371];
for i = 1:length(k_n)
y_obs = (-(cos(k_n(i) * L)- cosh(k_n(i) * L)) ...
* (sinh(k_n(i) * x_obs) + sin(k_n(i) * x_obs)) ...
+ (sin(k_n(i) * L)-sinh(k_n(i) * L)) ...
* (cosh(k_n(i) * x_obs) + cos(k_n(i) * x_obs))) ...
/(-(cos(k_n(i) * L)- cosh(k_n(i) * L)));
chap_3_plot( x_obs,y_obs)
end
plot(x_obs,zeros(1,length(x_obs)),'-.k','Linewidth',1);
leg1 = legend({'$Mode~1$','$Mode~2$','$Mode~3$','$Mode~4$'});
set(leg1,'Location','northeastoutside','Interpreter','latex');
set(leg1,'FontSize',16);
%%%%%%%%%%%% Function(s) %%%%%%%%%%%%
function chap_3_plot( x_obs,y_obs)
y_label = '\hat{\bar{w}}(x)';
nAme = ['$',num2str(y_label),'$'];
```

```
set(gca,'TickLabelInterpreter','latex')
set(gca,'fontsize',16)
min_x=0;
max_x=1;
grid on
grid minor
box on
hold on
plot(x_obs,y_obs,'Linewidth',2);
xlim([min_x,max_x])
ylim([-2,2])
xlabel('$\hat{x}$','Interpreter','Latex')
ylabel(nAme,'Interpreter','Latex')
end
```

B.2 Chapter 4

MATLAB®

```
close all; clear; clc;
summary = xlsread('summary_feature.xlsx','B2:J18');
x_obs = summary(:,end);
y_obs = summary(:,1:end-1);
%% Plot
for i = 1:8
figure(i)
plot_feature( x_obs,y_obs,i )
end
%%%%%%%%%%%%% Function(s) %%%%%%%%%%%%%
function plot_feature( x_obs,y_obs,i )
y_label = {'Peak\ge60\%', 'Peak\ge20\%','L2' ...
,'CL','KU','COR','mean~peak','frequency'};
nAme = ['$',num2str(y_label{i}),'$'];
set(gca,'TickLabelInterpreter','latex')
set(gca,'fontsize',16)
min_x=0;
max_x=50;
grid on
grid minor
box on
hold on
```

```
plot(x_obs,y_obs(:,i),'Linewidth',2);
p = scatter(x_obs,y_obs(:,i),40,'magenta','filled');
xlim([min_x,max_x])
xlabel('$x$','Interpreter','Latex')
ylabel(nAme,'Interpreter','Latex')
legend(p,{'$Observations$'},'Location' ...
,'southeast','Orientation','horizontal','Interpreter','Latex')
end
```

B.3 Chapter 5

MATLAB®

```
clear;close all;clc;
%% Data
summary = xlsread('summary_feature.xlsx','B2:J18');
x_obs = summary(:,end);%2:end-1 x_obs = y_obs/1000;
y_obs = summary(:,1);
vec =[1,2,3,4,5,6,7,8,9,10,11,12,13,15,16,17];% [1:17];%[1,2,3,4,5,6,7 ...
% ,8,9,10,11,12,13,15,16,17];% [1,4,6,8,10,14,17]; [1,3,9,15,17]; ...
% [1,2,3,4,14,15,16,17]; [1,2,3,4,5,6,17]
%vec = [4,6,8,10,14];
x_obs = x_obs(vec);
y_obs = y_obs(vec);

s_V=1E-6; %measurement noise std~0
%% GPML
l_01=5;%0.31; %initial guess for l
sG_0=1;
% covfunc= {'covSEard'}; %square exponential fct.
covfunc= {'covSEiso'}; %square exponential fct.
meanfunc = @meanConst;
likfunc = @likGauss; %Gaussian likelihood fct
hyp.mean =0;
%% Parameter Estimation
x_train = x_obs;
y_train = y_obs;
% Parameter initialization
ell = log([l_01]);sg=log(sG_0);
hyp.cov = [ell;sg]; %log(hyper-param.)
hyp.lik = log(s_V); %log(lik. hyper-parameters)
```

```
% Parameter estimation
[hyp,fx] = minimize(hyp, @gp,-10000,@infExact,meanfunc, covfunc   ...
,likfunc ,x_train ,y_train);

%% Optimal parameter values
l_GPML=exp(hyp.cov(1)); %opt. value for l
s_G_GPML=exp(hyp.cov(2)); %opt. value for s_G
%% Prediction
x_test = linspace(0, 50, 200)';%new covariates
[~,~,GP_m,GP_s2] = gp(hyp,@infExact, meanfunc,covfunc,likfunc   ...
,x_obs ,y_obs,x_test);
GP_m; %prediction means
GP_s2; %prediction variances

%% Negative log marginal likelihood
[nlZ, ~] = gp(hyp,@infExact, meanfunc,covfunc,likfunc ,x_obs ,y_obs);
%% Plot
hold on
% plot(x_obs,x_obs);
% scatter(x_obs,GP_m,'r');
f = [GP_m+2*sqrt(GP_s2); flipdim(GP_m-2*sqrt(GP_s2),1)];
fill([x_test; flipdim(x_test,1)], f, [7 7 7]/8,'LineStyle','none')
plot(x_test, GP_m); plot(x_obs, y_obs, '+','markers',12)
alpha(0.7)
grid on
grid minor
box on
leg1 = legend({'$Uncertainty,~\pm\sigma_{*|\mathcal{D}}$' ...
,'$Predictive~Mean,~{m}_{*|\mathcal{D}}$' ...
,'$Observation,~\mathcal{D}$'});

set(leg1,'Location','south','Interpreter','latex');
set(gca,'TickLabelInterpreter','latex')
set(0,'DefaultAxesFontName', 'Helvetica')
set(gca,'fontsize',16)
xlabel('$Pipe~Location~(m)$','Interpreter','Latex')
ylabel('${m}_{*|\mathcal{D}}$','Interpreter','Latex')
%% Save file
nAme = ['GP_2','.pdf'];
set(gcf,'Units','inches');
```

```
screenposition = get(gcf,'Position');
set(gcf,...
'PaperPosition',[0 0 12 4],...   %12 4
'PaperSize',[12 4]);
saveas(gca,nAme);
nAme_2 = ['pdfcrop', nAme, nAme];
system(nAme_2);
```