

Carnegie Mellon University

CARNEGIE INSTITUTE OF TECHNOLOGY

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF Doctor of Philosophy

TITLE A Data-Driven Framework Based on Sparse Representation of
Ultrasonic Guided-Waves for Online Damage Detection of Pipelines

PRESENTED BY Matineh Eyb Poosh

ACCEPTED BY THE DEPARTMENTS OF

Civil and Environmental Engineering

Mario Berges August 13, 2015
CO-ADVISOR, MAJOR PROFESSOR DATE

Haeyoung Noh August 17, 2015
CO-ADVISOR, MAJOR PROFESSOR DATE

David A. Dzombak August 13, 2015
DEPARTMENT HEAD DATE

APPROVED BY THE COLLEGE COUNCIL

Vijayakumar Bhagavatula August 24, 2015
DEAN DATE

**A Data-Driven Framework Based on Sparse Representation of
Ultrasonic Guided-Waves for Online Damage Detection of Pipelines**

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy
in
Civil and Environmental Engineering

Matineh Eyb Poosh

B.S., Civil Engineering, Tabriz University, Iran

M.S., Construction Engineering and Management,
Middle East Technical University, Turkey

M.S., Advanced Infrastructure Systems,
Carnegie Mellon University, USA

Carnegie Mellon University
Pittsburgh, PA
July, 2015

To Mother Earth.

Acknowledgement

This thesis would not have been possible without the guidance and support of many people who accompanied me throughout my Ph.D. I would like to thank my adviser Dr. Mario Berges, who never withheld his support. He indeed has had an immeasurable contribution in my journey in CMU, both academically and personally. His inquisitive and curious attitude developed a researcher's approach in me, which I have no doubt I will benefit from throughout my life. I would like to thank him for his friendship, unconditional guidance (even if it takes him to come to the lab on a Saturday to help me fixing a broken equipment), and his efforts in making the Ph.D. an enjoyable process for his students.

I would like to thank my co-adviser Dr. Haeyoung Noh, who gladly agreed to be part of this journey. This thesis is indebted to her guidance, challenging questions (answering each of which was a step forward in my thesis), and valuable insights. She always encouraged me, and was always there when I needed her.

I am grateful to my committee members, Dr. Irving Oppenheim and Dr. Joel Harley for their time and insightful discussions. I feel fortunate to have the privilege of learning from them throughout my Ph.D.

Other people I owe thanks to are Dr. David Greve and Dr. Warren Junker for their valuable feedback on my work, and Dr. Chang Liu for providing the field data that I used for part of the experiments in this dissertation.

I would also like to express my sincere gratitude to Dr. James H. Garrett. Regardless of how busy he was, he never stopped caring. His unconditional support and encouragement have had a significant impact on my experience in CMU.

I would also like to acknowledge Civil and Environmental Engineering department at Carnegie Mellon University, and Julia and Michael Ellegood research fellowship for supporting this work. I am also grateful to kWantera Inc. for their generous contribution in my semester at CMU.

I would like to thank the best lab mates of all time, Insoo, Leneve, Emre, Suman, and Jingkun.

I gained incredible friends in Pittsburgh, without whom this Ph.D. would have been very difficult and much less enjoyable. Emre, thank you for going through every ups and downs of this path with me, and never giving up on being a real friend. Sasan, thanks for always (always) being there for me. Your ability in making me laugh aloud at any time deserves a Nobel Prize. Laleh, Shadi and Azarin, thank you for being the girlfriends that every Ph.D. student would need. I cannot imagine finishing my Ph.D. without the “girls’ night out”.

I would also like to thank Julian Lamy for his love and priceless companionship. Julian, you taught me to be adventurous, ambitious, and not settle for less than what I can be truly proud of. I feel very lucky to have you.

Lastly, I would like to express my thanks to my parents; Susan and Hamid, my sister; Sana, and my brother; Hosein. Thank you for your never-ending support. Thank you for being on the other side of the phone whenever I needed, no matter what time it is there. What I have achieved would not have been possible without any of you.

Abstract

This work addresses some of the open challenges in guided-wave based structural health monitoring (SHM) of pipelines. In this dissertation, we review these challenges under three headings: (a) Multiple modes, (b) Multi-path reflections, and (c) Sensitivity to environmental and operational conditions (EOCs). The objective is to develop damage detection methods that (1) simplify guided-wave signals, and (2) have low sensitivity to EOC variations.

First, I propose a supervised method for online damage detection. The detection performance is maximized under variety of EOCs, by imposing a sparsity constraint on the signals. In the training stage, data is recorded from an intact pipe, as well as the pipe with structural abnormalities. During the monitoring stage, test signals are projected on the extracted sparse discriminant vector, and these projections are used as damage-sensitive features. I conduct a diverse set of laboratory and field experiments to investigate and to validate the extent to which EOC variations, as well as the differences in characteristics of the structural abnormality in training and test data, can influence the discriminatory power of the damage-sensitive features. The validation results suggest that a simple binary-labeled training data (*i.e.*, undamaged/damaged), obtained under a limited range of EOCs, is sufficient for the proposed method. In other words, the detection method does not require prior knowledge about the characteristics of the damage (*e.g.*, size, type, location), and/or a training dataset that is obtained from a wide range of EOCs.

Next, I propose an unsupervised method to address some of the limitations of the aforementioned supervised approach. The unsupervised approach eliminates the need for training data captured from a pipe with structural abnormality, which can be a challenge for some applications such as pipes with restricted accessibility. Therefore, the damage-related training parameters that may affect the detection performance of the supervised method are not an issue for the unsupervised approach. The proposed unsupervised method takes advantage of two facts that are further verified throughout this work: (1) high-energy arrivals are less sensitive to EOC variations compared to the rest of the signal, and (2) damage changes the energy-content and/or time-location of high-energy arrivals in the signal. For this method, the training dataset is not labeled, but is assumed to mostly include intact signals. In the training stage, a sparse subset of high-energy arrivals from intact pipes is extracted so that energy estimation error is minimized. My experimental analysis proved that high-energy arrivals in intact signals are located at different time-points than those in damaged signals. Therefore, using the extracted sparse subset (which is mainly composed of high-energy arrivals), the energy estimation error will increase as damage occurs. The proposed method is proved successful for online detection of damage under varying EOCs. It is also shown that the wider the range of EOCs in the training dataset, the better the detection performance. This range, however, is not required to be comprehensive of all possible testing scenarios. For example, for a test dataset in which temperature varies between 24°C and 39°C, a training dataset with temperatures ranging between 24°C and 30°C results in separation accuracy of 99%, and detection delay of five observations, captured in one-minute intervals.

Content

Acknowledgements.....	v
Abstract.....	viii
List of Figures.....	xiii
List of Tables.....	xvi
Chapter 1. Introduction and Problem Statement.....	1
1.1 Background.....	1
1.2 Guided-waves and their Advantages for Structural Health Monitoring (SHM) of Pipelines.....	3
1.3 Challenges for Guided-wave based SHM.....	6
<i>1.3.1 Multiple Modes.....</i>	<i>6</i>
<i>1.3.2 Multi-path Reflections.....</i>	<i>8</i>
<i>1.3.3 Sensitivity to Environmental and Operational Conditions (EOCs).....</i>	<i>10</i>
Chapter 3. Research Objective and Scope.....	12
2.1 Research Objective.....	12
2.2 Research Scope.....	13
Chapter 3. Experimental Setup.....	16
3.1 Aluminum Pipe with Mass Scatterer.....	16
3.2 Steel Pipe with Crack and Mass Loss.....	19
3.3 Steel Pipe Operating Under Multiple Varying EOCs.....	20
Chapter 4. A Supervised Approach for Damage Detection of Pipelines.....	22
4.1 Introduction.....	22
4.2 Motivation.....	23
4.3 Method.....	25

4.3.1 Research Questions.....	26
4.3.2 Overview of the Sparse Discriminant (SD) Method.....	26
4.3.3 Evaluation Criteria for the SD Method.....	31
4.4 Experimental Investigation of the SD Method.....	33
4.4.1 Temperature Variation.....	34
4.4.2 Damage Size and Temperature Variation.....	40
4.4.3 Damage Location and Temperature Variation.....	44
4.4.4 Damage Size, Damage Location, and Temperature Variation.....	50
4.4.5 Physical Intuition about the SD Coefficients.....	51
4.5 Validation of the SD Method.....	55
4.5.1 Validate the Sparsity Assumption with Field Data.....	55
4.5.2 Online Damage Detection under Multiple EOCs.....	57
4.5.3 Detection of a Small Crack and Mass Loss.....	59
4.5.4 Non-Gaussianity of Guided-waves.....	62
Chapter 5. An Unsupervised Approach for Damage Detection of Pipelines.....	63
5.1 Introduction.....	63
5.2 Motivation.....	64
5.3 Method.....	66
5.3.1 Research Questions.....	67
5.3.2 Overview of the Sparse Energy (SE) Method.....	68
5.4 Experimental Investigation of the SE Method.....	72
5.4.1 Sparse Representation of Signals for Energy Estimation (Question #1)....	72
5.4.2 Stability of the SE Method.....	76
5.4.3 Sensitivity of the Sparse Signals to Temperature Variation (Question #3).	76
5.4.4 Damage Detection with SE Method (Question #2).....	79
5.4.5 Sensitivity of Damage-sensitive Features (ΔE s) to Temperature Variation (Question #3).....	83
5.4.6 Physical Intuition about the SE Coefficients.....	86
5.4.7 Selecting Regularization Scalar (ξ) for SE Method (Question #1).....	91

5.5 Validation of the SE Method.....	95
5.5.1 Methods for Evaluating the Classification and Detection Power of ΔE s...	96
5.5.2 Damage Detection with Varying Training and Test Temperatures.....	99
5.5.3 Detection of a Small Crack and Mass Loss	101
5.5.4 Online Damage Detection under Multiple EOCs.....	103
5.5.5 Effects of Including Damaged Signals in the Training Dataset.....	105
Chapter 6. Conclusion.....	108
6.1 Summary of the Work.....	108
6.2 Contributions.....	112
6.3 Limitations.....	116
6.4 Future Work.....	117
Bibliography.....	119

List of Figures

Figure 1.1: Map of the major natural gas and oil pipelines in United States (PHMSA, 2012).....	2
Figure 1.2: Guided-wave modes and corresponding displacements in cylindrical media.....	5
Figure 3.1: Laboratory setup for controlling temperature variation: 1) Thermally insulated box to contain the experimental pipe segments, 2) Thermostat, 3a & 3b) PZT transducer and receiver, respectively, 4) Grease-coupled mass Scatterer to simulate structural abnormality.....	17
Figure 3.2: Unrolled view of the laboratory aluminum pipe segment, for different transducer-mass layout used throughout this work. Note: The three symbols in the layout names (1 vs. 2, Loc1 vs. Loc2, and big vs. sml) indicate the transducer setup, mass location, and mass size, respectively.....	18
Figure 3.3: The steel pipe used to examine the application of the proposed methods for detection of multiple actual damages, namely a crack and mass loss.....	19
Figure 4.1. A schematic unrolled view of a pipe, illustrating example wave propagation paths passing through undamaged (dashed arrows) and damaged (solid arrows) sections of a pipe.....	24
Figure 4.2: Application of the proposed SD method for continuous monitoring of pipelines.....	27
Figure 4.3: (a) Average detection accuracy of 2-fold CV, for different training/test temperature combinations ($24^{\circ}\text{C} \leq T \leq 38^{\circ}\text{C}$), using the class labels predicted by SD method (\hat{Y} s) as the only damage-sensitive feature for clustering. (b) ROC-curves for training/test scenarios with $0^{\circ}\text{C} \leq \Delta T \leq 8^{\circ}\text{C}$. T_{tr} : Training temperature, T_{tst} : Test temperature.....	35
Figure 4.4: Variation of the distance between predicted labels for intact and damaged test observations at different training/test temperature scenarios. T_{tr} : Training temperature, T_{tst} : Test temperature, \bar{Y}_{int} : Average predicted labels for intact test observations, \bar{Y}_{dmg} : Average predicted labels for damaged test observations, $\sigma_{\hat{Y}_{int}}$:	

Standard deviation of the predicted labels for intact test observations.....	37
Figure 4.5: Illustration of online damage detection with varying temperature, using class labels predicted by SD method. (a) Temperature difference between the training dataset and the monitored observations, (b) Correlation between the monitored observations and average of training intact signals, (c) Predicted class labels for the monitored observations using equation 4.3.....	38
Figure 4.6: Mean of the coefficients at all sample points in time, normalized with the standard deviation of the magnitudes (equation 4.10), for $24^{\circ}\text{C} \leq T_{tr} \leq 32^{\circ}\text{C}$. Note: First arrival times are calculated for 2-Loc2-big pipe using PCDisp (Seco and Jiménez, 2012).....	53
Figure 4.7: Cross-correlation between the mean of the coefficients trained with small damage (2-Loc2-sml) and big damage (2-Loc2-big), at $24^{\circ}\text{C} \leq T_{tr} \leq 33^{\circ}\text{C}$	54
Figure 4.8: Cross-correlation between the mean of the coefficients trained for two different locations of damage (2-Loc1-big and 2-Loc2-big), at $24^{\circ}\text{C} \leq T_{tr} \leq 33^{\circ}\text{C}$	55
Figure 4.9: Variation of sparsity ratio (solid red line) and detection accuracy of the SVD method (Liu et al., 2013) (dashed black line) for different values of regularization scalar ξ in equation 4.2.....	56
Figure 4.10: Online damage detection using \hat{Y} s for filed data explained in Section 3.3: (a) Temperature variation of the test observations, (b) Fluid flow rate variation of the test observations, (c) Predicted class labels (\hat{Y} s) for test observations.....	58
Figure 4.11: Online detection of a 2.5 cm long oblique crack (with maximum thickness of 0.34 mm and maximum depth of 0.7 mm), and a subsequent small mass loss (with maximum depth of 0.25 mm), in a Schedule-40 steel pipe.....	61
Figure 5.1: A schematic unrolled view of a pipe, illustrating example wave propagation paths in an undamaged pipe.....	66
Figure 5.2: Application of the SE method for continuous monitoring of pipelines.....	69
Figure 5.3: Variation of (a) average training error (equation 5.6), (b) sparsity ratio (equation 5.7), and (c) ℓ_1 norm of \vec{X} vector, with respect to regularization scalar ξ . Note: Regularization scalars are presented in logarithmic scale.....	75
Figure 5.4: (a) Correlation between a reference complete/sparse signal ($T = 24^{\circ}\text{C}$) and complete/sparse test signals at different temperatures, for different values of ξ . (b) Deviation range for the correlations corresponding to $\xi = 0$ and $\xi = 1.0$. Note: Error bars represent 2σ	77
Figure 5.5: (a) The coefficients obtained by applying the SE method on intact signals, (b) the coefficients obtained by applying the SE method on damaged signals, (c) the	

difference between (a) and (b), <i>i.e.</i> , $X^{int} - X^{dmg}$	81
Figure 5.6: (a) The coefficients obtained by applying SE method on damaged signals, (b) the coefficients obtained by applying SD method on a training dataset including both intact and damaged signals.....	83
Figure 5.7: Correlation of energy estimation errors (ΔE s) of (a) intact test signals and (b) damaged test signals, for different widths of training temperature range (δT). Note: δT is the difference between the maximum and minimum temperatures within different training datasets. μ is the mean of the correlations at each δT , and σ is the standard deviation of the correlation values at each δT	86
Figure 5.8: Normalized coefficients obtained by (a) SE method using intact signals, (b) SE method using damaged signals, and (c) SD method using both intact and damaged signals. Note: the aluminum pipe layout <i>2-loc2-sml</i> shown in Figure 3.2d is used for these experiments.....	88
Figure 5.9: (a) Average training error at different values of regularization scalar ξ (see equation 5.6), (b) Correlation between the energy estimation errors ($\Delta E_j \forall j \in \{1, \dots, m_{tr}\}$, see equation 5.5) and the temperatures of training signals (T_j) at different values of regularization scalar ξ . E_j^{act} : energy of the complete training signals.....	92
Figure 5.10: (a) Average training error at different values of regularization scalar ξ (see equation 5.6), (b) Correlation between the energy estimation errors ($\Delta E_j \forall j \in \{1, \dots, m_{tr}\}$, see equation 5.5) and the temperatures of training signals (T_j) at different values of regularization scalar ξ . E_j^{act} : energy of the complete training signals.....	94
Figure 5.11: Change in the percentage of nonzero coefficients for different levels of sparsity (<i>i.e.</i> , different values of regularization scalar ξ) corresponding to (a) first set of arrivals, (b) second set of arrivals, and (c) later arrivals (see Figure 5.8 for clarification on first, second and later set of arrivals).....	95
Figure 5.12: Online detection of a 2.5 <i>cm</i> long oblique crack (with maximum thickness of 0.34 <i>mm</i> and maximum depth of 0.7 <i>mm</i>), and a subsequent small mass loss (with maximum depth of 0.25 <i>mm</i>), in a Schedule-40 steel pipe.....	103
Figure 5.13: An example case from field data where perfect classification of the ΔE s with K-means and SVM based approaches is not be possible, but occurrence of damage can be detected using KL based approach.....	105
Figure 5.14: (a) Classification accuracies, and (b) delays in detection of occurrence of damage, corresponding to different ratios of damaged signals in the training dataset....	106
Figure 6.1: Overview of the proposed online damage detection framework.....	111

List of Tables

Table 4.1: Average statistics of 2-fold CV, reflecting the detection performance of class labels predicted by SD method (\hat{Y} s), for a total of 3,025 training/test scenarios given in Figure 4.3.....	36
Table 4.2: Detection performance of \hat{Y} s for online monitoring of pipes, under varying temperatures. Note: <i>1-Loc1-sml</i> layout is used. Test data includes 610 intact and 610 damaged observations, recorded in 1-minute intervals.....	39
Table 4.3: Sparsity ratio (Sr., equation 4.4) statistics as a result of training the SD algorithm with datasets of different temperatures.....	40
Table 4.4: Average detection statistics for a total of 1,250 training/test scenarios with different damage sizes and temperatures. Note: $25^{\circ}\text{C} \leq T_{tr} \text{ \& } T_{tst} \leq 33^{\circ}\text{C}$. Pipe layout <i>1-Loc1-sml</i> is used for training and layout <i>1-Loc1-big</i> is used for testing.....	42
Table 4.5: Detection performance of \hat{Y} s for online monitoring of pipes, when sizes of damage, as well as the temperatures of training and test datasets are different. Note: <i>1-Loc1-sml</i> pipe layout is used for training, and both <i>1-Loc1-sml</i> and <i>1-Loc1-big</i> layouts are used for testing. Observations are recorded in 1-minute intervals.....	43
Table 4.6: Sparsity ratio (Sr, equation 4.4) statistics as a result of training SD algorithm with signals from a pipe introduced to two different sizes of abnormality, at different temperatures ($24^{\circ}\text{C} \leq T_{tr} \leq 38^{\circ}\text{C}$).....	44
Table 4.7: Average detection statistics for 900 training/test combinations trained and tested with different damage locations, and temperatures. Note: $25^{\circ}\text{C} \leq T_{tr} \text{ \& } T_{tst} \leq 33^{\circ}\text{C}$. Pipe layout <i>2-Loc1-big</i> is used for training, and layout <i>2-Loc2-big</i> is used for testing.....	48
Table 4.8: Detection performance for online monitoring of pipes, using \hat{Y} s, when temperature, as well as the location of damage in training and test data is different. Note: <i>2-Loc1-big</i> pipe layout is used for training, and <i>2-Loc2-big</i> layout is used for testing. Observations are recorded in 1-minute intervals.....	49
Table 4.9: Sparsity ratio (Sr, equation 4.4) statistics as a result of training SD algorithm with signals from a pipe with damage at two different locations, at different	

temperatures ($25^{\circ}\text{C} \leq T_{tr} \leq 33^{\circ}\text{C}$).....	50
Table 4.10: Detection performance of \hat{Y} s for online monitoring of pipes, when temperature, damage size, and damage location in training and test data is different. Note: <i>2-Loc2-big</i> is used for training, and <i>2-Loc1-sml</i> is used for testing.....	51
Table 4.11: Average online detection performance and sparsity ratios, for 10-fold CV, using field data explained in Section 3.3, for four summer days and three winter days. Observations are recorded in 1-minute intervals.....	59
Table 4.12: Percentage of the sample points, after the first arrivals, for which the null hypothesis of Gaussianity is rejected at 5% significance level.....	62
Table 5.1: The breakdown of the total energy, and the overlap between the coefficients extracted by SE method, using intact and damaged signals separately. Note: The aluminum pipe layout <i>2-loc2-sml</i> shown in Figure 3.2d is used for these analyses.....	88
Table 5.2: The breakdown of the total energy of the sparse signals, and the overlap between the coefficients extracted by SE method, using intact and damaged signals separately. Note: The aluminum pipe layout <i>2-loc2-sml</i> shown in Figure 3.2d is used for these analyses.....	90
Table 5.3: Average detection statistics for 10-fold cross validation with the ΔE values calculated with the coefficients corresponding to different ranges of training temperature. Note: The temperatures of test signals vary between 24°C and 39°C	100
Table 5.4: Average detection statistics for 10-fold cross validation with the ΔE values calculated for the intact steel pipe, the pipe with a small crack, and the pipe with both crack and small mass loss.....	102
Table 5.5: Average detection statistics for 10-fold cross validation with the ΔE values calculated for a hot-water supply piping system, with and without a structural abnormality.....	104

Chapter 1

Introduction and Problem Statement

1.1 Background

Pipelines are crucial infrastructure components in several applications at different scales, ranging from building-level piping systems to inter-state transmission pipelines. According to the US Department of Transportation's report (DOT, 2010), 100% of the natural gas and 71% of oil and refined petroleum products consumed in the United States are transferred by pipelines (see Figure 1.1). Reliable prediction of degradation in structural integrity of pipes throughout their lifetime is important to ensure delivery of expected services, to decrease environmental/human-safety risks associated with not/late-detected damages, and to reduce cost/time of repair and handling impacts of damages.

According to the Department of Transportation, Pipeline and Hazardous Materials Safety Administration report (PHMSA, Pipeline Incident 20 Year Trend), 10,613 incidents have been reported in US between 1994 and 2013, which translates to an average annual rate of 531

incidents, 19 fatalities, 71 injuries, property damage of more than \$308.8 million, and more than 133,000 barrels of spilled hazardous liquid, including crude oil. Recent advances in sensing and computing, however, do not seem to have had substantial impact on improving these statistics. For example, only 28% of the reported incidents in the instrumented pipes were identified by data acquisition and testing systems between 2010 and 2012 (PHMSA, 2012). According to a study conducted in 2013 at National Oceanic and Atmospheric Administration (NOAA) and the University of Colorado (Karion et al., 2013), the ratio of the leakage to the average hourly natural gas production can reach up to 14% in certain locations. Another example of pipeline application affected by missed or delayed detection of damages is water transmission. As reported by the American Water Works Association (AWWA), 237,600 breaks per year occur in the US, which leads to an approximate \$2.8 million lost in annual revenue (EPA, 2007). One of the main causes of these losses is leakage due to different types of damages in pipes (Shinozuka et al., 2010).

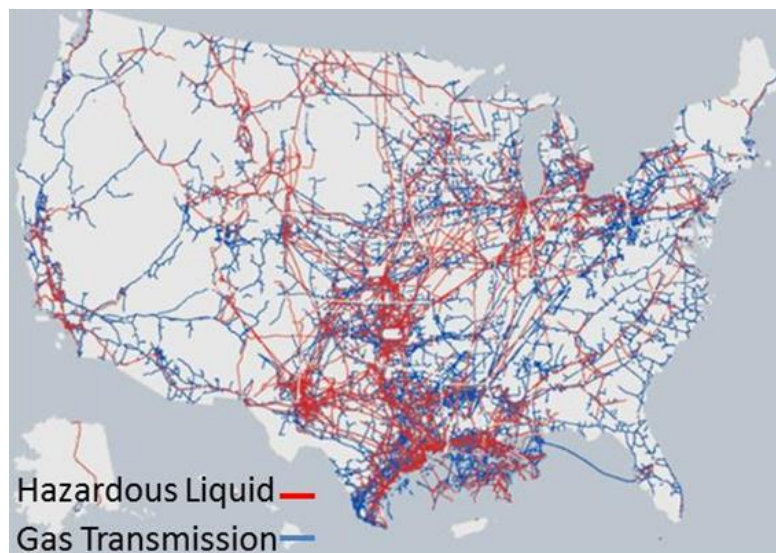


Figure 1.1: Map of the major natural gas and oil pipelines in United States (PHMSA, 2012)

Data acquisition and sensing systems, however, are valuable for continual monitoring, and for cases where accessing pipeline is not feasible (PHMSA, 2012). Therefore, research is needed to improve the effectiveness of these systems.

1.2 Guided-waves and their Advantages for Structural Health Monitoring (SHM) of Pipelines

Difficulties, cost, and safety risks associated with accessing different portions of the pipes make non-destructive evaluation (NDE) techniques attractive for pipeline damage detection and monitoring (Cawley et al. 2003). In the majority of applications, large portions of the pipelines are either insulated or covered by other structures or materials such as soil (Alleyne et al. 2001). Thus, the reliability of traditional destructive methods for timely detection of damages is limited, as only observable/accessible damages may be identified with these methods.

Among various techniques, the benefits of guided-waves for non-destructive evaluation (NDE) purposes have been discussed for over half a century, since Worlton recognized their potential (Worlton, 1957). Guided-waves are mechanical stress waves propagating along a media, guided by its boundaries. There are different types of guided-waves depending on the type of the structure and excitation conditions, but they all have one requirement in common: a well-defined boundary. If the thickness of the wave guide (here, a cylindrical media, *i.e.*, a pipe) is comparable to the wavelength of the excited wave modes, guided-waves will travel along the

axial, radial, and circumferential direction of the pipe wall (see Figure 1.2). Therefore, generally, three types of modes can be excited in a pipe-like structure:

Longitudinal modes (L): travel throughout the axial direction of the pipe, and are identifiable with axial displacements (u_z) and compressional displacements along the radial direction (u_r). These modes are axisymmetric (see Figure 1.2).

Torsional modes (T): travel parallel to the circumferential direction (θ) and introduce shear motions resulting in axial displacements (u_z) and radial displacements (u_r). These modes are axisymmetric (see Figure 1.2).

Flexural modes (F): are not axisymmetric and introduce displacements in circumferential direction (u_θ) in addition to u_z and u_r (see Figure 1.2).

Equations 1.1-1.3 summarize the particle displacements for a cylindrical media as given by Rose (2004):

$$u_r = U_r(r) \cos(n\theta) \cos(\omega t + kz) \quad (1.1)$$

$$u_\theta = U_\theta(r) \sin(n\theta) \cos(\omega t + kz) \quad (1.2)$$

$$u_z = U_z(r) \cos(n\theta) \sin(\omega t + kz) \quad (1.3)$$

In these equations, $n \in \{0,1,2, \dots\}$ is the circumferential order, for different wave orders ($m \in \{0,1,2, \dots\}$) of the longitudinal, torsional and flexural modes.

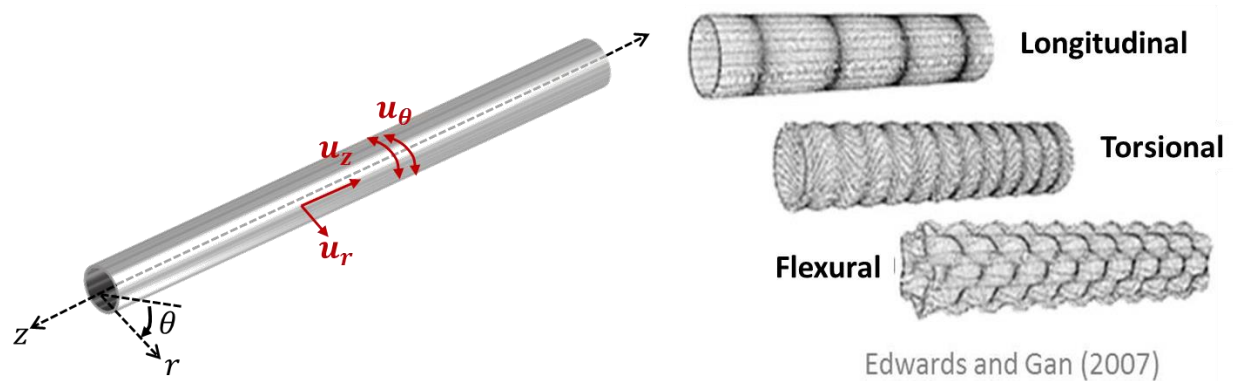


Figure 1.2: Guided-wave modes and corresponding displacements in cylindrical media

During the past decade, non-destructive monitoring and evaluation techniques for pipelines using guided-waves have been widely considered by researchers and service providers in different applications ranging from water supply pipes to gas/oil transmission, chemical or power generation pipes (*e.g.*, Alleyne et al., 2001; Siqueira et al. 2004; Sun et al. 2000; Wang et al. 2010). This is due to many favorable characteristics of guided-waves compared to conventional methods and it makes these systems a promising candidate for continuous and automated damage detection of pipes (Rose, 2004; Vanlanduit et al., 2005). The advantages include full coverage of the thickness and surface of the pipe, long travel distances without significant energy loss, and high sensitivity to different sizes/types of damage. In addition, guided-wave based SHM systems can operate with small number of low-cost, low-power, transducers, making their implementation efficient (Cawley et al., 2003; Davies et al., 2008; Demma et al., 2003; Rose, 2004; Shull, 2002; Vanlanduit et al., 2005).

1.3 Challenges for Guided-wave based SHM

Despite their many advantages, real-world application of guided-waves for pipeline SHM is still quite limited (Cawley et al., 2003). The challenges leading to this can be discussed under three headings: (a) Multiple modes, (b) Multi-path reflections, and (c) Sensitivity to environmental and operational conditions (EOCs).

1.3.1 Multiple Modes

Generally, three types of modes can be excited in a pipe-like structure: (a) longitudinal modes, $L(0,m)$, (b) torsional modes, $T(0,m)$, and (c) flexural modes, $F(n,m)$, m and n referring to wave orders and circumferential order of modes, respectively (Rose, 2004). The presence of at least two modes at any given frequency is one of the factors making the guided-wave problems complex as compared to bulk waves (Demma et al., 2003). Another complexity arises from the dispersive nature of guided-waves (Rose, 2004). That is, propagation velocities of different wave modes are themselves functions of frequency.

To address these challenges, one of the most widely considered solutions is to excite limited number of modes in a non/less dispersive range by using either an array of transducers or multi-element transducers (*e.g.*, Davies et al., 2008; Demma et al., 2003). Among all, axisymmetric modes, $L(0,2)$ and $T(0,1)$ are mostly used in practice, mainly because they are easier to excite and the acoustic field is relatively simple (Alleyne et al., 2001; Demma et al., 2004, 2003). Studies such as Alleyne et al. (1998) have shown that reflection coefficients of incident $L(0,2)$

mode are linearly related to the circumferential extent of a damage. Wang et al. (2010) uses distinctive characteristics of the reflected $L(0,2)$ mode from the front and end edge of a damage to characterize its axial extent. Lowe et al. (1998a) investigated the reflection of the non-axisymmetric modes $F(1,3)$ and $F(2,3)$ as a result of mode conversion from a non-symmetric through-thickness notch when $L(0,2)$ mode is used. However, L modes can be vulnerable to additional reverberation and energy leakage at the interference of fluid inside/outside the pipe (Rose, 2004). Moreover, exciting $L(0,2)$ mode in the less dispersive range requires careful design of the transducer system to suppress $L(0,1)$ mode (Demma et al., 2004, 2003). These issues have motivated the use of $T(0,1)$, which has simpler dispersion characteristics (*i.e.*, non-dispersive in all frequencies), insensitivity to liquid interference, and thus, longer travel distance. $T(0,1)$ is currently the most frequently used mode for pipeline NDE (Løvstad and Cawley, 2012). Demma et al. (2003) have shown linear dependency between the reflection of incident $T(0,1)$ mode and radial depth of a damage. Nurmalia et al. (2013) relate the change in the thickness of the pipe to the change in the group velocity of the propagated $T(0,m)$ modes, which is the result of mode conversion from higher orders to the lower orders as the thickness is decreased due to damage.

As can be inferred from these studies, different wave modes are sensitive to different characteristics of the damage. Considering the fact that the geometry of a real damage in pipes can extend in any or all of the axial, radial and circumferential directions, damage diagnostics based on single-mode excitation may lose the benefit of multiple modes. Moreover, as some of these studies have shown, mode conversions can still occur when the incident mode interferes with damage or other material such as flowing fluid (Aristégui et al., 2001). That is, the excited non-dispersive mode can be converted to other mode(s) that may be dispersive. In addition, even

in the case of successful single-mode excitation, multi-path reflections from structural features and/or damage, as well as the EOC effects, which will be discussed in the next sections, cause the guided-waves, traveling in an operating pipe, to be the result of superposition of multiple modes (Lu and Michaels, 2005). Other challenges include implementation difficulties/costs compared to simple broad-band multi-mode excitation transducers (Demma et al., 2004; Nurmalia et al., 2013).

1.3.2 Multi-path Reflections

Reflections from the features of the structure (*e.g.*, boundaries, pipe welding, damage, *etc.*), and their superposition, adds to the complexity of guided-waves. A number of studies have worked on solutions to address the challenges of multi-path reflections. Many approaches rely on baseline-subtraction, (*e.g.*, Croxford et al., 2010, 2007). Ideally, baseline-subtraction will remove the background complexities that are due to structural features. However, such reference signals should be recorded under similar EOCs as the new signal, or the effects of the EOC variation need to be compensated. Among all, effects of temperature have been the most widely studied (Croxford et al., 2007; Lu and Michaels, 2005; Scalea and Salamone, 2008; Schulz et al., 2003). Temperature effects on wave velocity are typically approximated as stretching/compressing of the signal. However, such stretching methods may successfully approximate only small ranges of temperature variation, *i.e.*, 0.5°C to 1°C, depending on the complexity of the structure and the number of propagating modes (Clarke et al., 2009; Lu and Michaels, 2005). Moreover, other EOCs, such as fluid flow rate and inner pressure (Degtyar et al., 1996; Eybpoosh et al., 2014a, 2014b; Harley et al., 2012), can further degrade the performance of these methods.

A number of studies have employed methods based on time-frequency analysis to cope with the complexity of guided-waves. These applications include, among other approaches, denoising and rectifying damage-sensitive components extracted through wavelet transforms (Siqueira et al., 2004), characterization of damages through matching pursuit decomposition (Tse and Wang, 2013), deriving dispersion curves (Niethammer et al., 2000), extracting damage-sensitive features by tracking the principal frequency components through time (Guerrero-Mosquera et al., 2010), and blind source separation (Belouchrani and Amin, 1998). One inherent limitation of most of these methods is the *uncertainty principle*, which states that achieving high frequency resolution requires sacrificing time resolution, and vice versa (Cohen, 1989). This becomes even a bigger challenge in a multi-modal signal in which multiple modes with relatively similar characteristics are excited at a given frequency. Moreover, the majority of these methods strongly depend on the selection of parameters like mother wavelet, dictionary of atoms, scaling and shift variables, detection thresholds, *etc.* It is also notable that the sensitivity of such parameters to EOC variations remains to be incorporated into the methods.

Recently, Harley and Moura (Joel B. Harley and Moura, 2013) developed a method for recovering a denoised signal by removing random noise and multipath signal interference for lamb waves in plates. The method uses the sparse nature of the lamb waves in the frequency-wavenumber domain, and identifies the sparse solutions satisfying equations of motions, using signals obtained from a network of transducers. This method has shown promising results in localizing damage in a plate by comparing the monitored signals to the reference denoised signal using a coherent matched field processor (Harley and Moura, 2014; J.B. Harley and Moura, 2013; Joel B. Harley and Moura, 2013). However, further developments are still needed to

examine the performance of the method when the EOCs of test and reference signals are different, and when different reference signals are associated with different EOCs (*e.g.*, non-uniform temperature change in the plate). In addition, relying on a network of transducers can be a challenge for certain NDE applications, since it calls for wider spatial access to the structure. Finally, yet importantly, the disregarded multi-path reflections may include information regarding damage, as will be illustrated later in this manuscript. The findings of my work suggest that denoising multi-modal multi-path guided-waves, for damage diagnosis, should not simply mean removing the multi-path reflections, but rather retaining damage information from the whole signal.

1.3.3 Sensitivity to Environmental and Operational Conditions (EOCs)

Guided-wave based damage detection of pipelines becomes even more challenging when EOCs vary over time, *e.g.*, changes in temperature, flow rate of the fluid carried by the pipe, inner pressure, interference with coupling material, *etc.*, (Croxford et al., 2010; Eybpoosh et al., 2015, 2014b; Harley and Moura, 2014; Harley and Moura, 2013; Liu et al., 2012b; Schulz et al., 2003). EOC effects range from generation of additional modes to changes in the wave velocity, changes in attenuation rate, shape distortion, and so on, (Aristégui et al., 2001; Long et al., 2003a; Scalea and Salamone, 2008). Such effects degrade the performance of damage diagnosis, by masking and/or appearing as the changes caused by structural anomalies and introducing type I and II errors.

Many studies have investigated, both theoretically and experimentally, the effects of EOCs, such as temperature and coupling material, on guided-waves (Aristégui et al., 2001; Scalea and Salamone, 2008; Schulz et al., 2003). However, few studies have tried to incorporate such effects into damage diagnosis approaches. These studies mainly include stretching of the reference signals to incorporate the effects of temperature on wave velocity of guided waves into baseline-subtraction methods, *e.g.*, (Croxford et al., 2007; Lu and Michaels, 2005). The challenges and limitations of such methods are briefly discussed in Section 1.3.2, and further illustrated in one of the authors' works (Eybpoosh et al., 2015). The missing physical and analytical intuition about the way EOCs affect different aspects of damage diagnostics, limits their extensibility to diverse operating conditions.

During the past two decades, a number of researchers have utilized data analysis techniques to overcome some of the EOC challenges and enhance the application of guided-waves for damage diagnosis, *e.g.*, (Harley et al., 2012; Liu et al., 2012a; Lu and Michaels, 2009; Ying et al., 2013). Statistical and signal processing techniques have shown high potential for extracting damage-sensitive features that are less sensitive to the particular EOCs considered. However, improvements are still needed to address limitations such as dependence on a network of transducers, reliance on a dictionary of atoms, case-specific tuning parameters, and linear decomposition of multi-modal signals whose bases may be related non-linearly (for more detailed discussion on this topic please refer to the authors' work (Eybpoosh et al., 2015)).

Chapter 2

Research Objective and Scope

2.1 Research Objective

The complexity of guided-waves is rooted in three features discussed in Section 1.3. The goal is to overcome these challenges for damage detection of pipes, while addressing the limitations of the current approaches. That is, the objective of this thesis is to develop feature extraction and damage detection methods that (a) simplify guided-wave signals, without the need for prior knowledge about the damage characteristics (*e.g.*, type, size, location), and (b) have low sensitivity to EOC variations, so that they can be extensible to diverse operation scenarios. Next section summarizes the scope within which this objective is pursued.

2.2 Research Scope

Before moving forward, it is important to define the scope within which the objective of this work will be satisfied.

Diagnosis Task

Among all diagnosis functionalities (*e.g.*, characterization, localization, *etc.*), this work is focused on damage detection. As will be shown later in this manuscript, the proposed methods have shown potential for identifying the change in the severity of damage. However, this work does not explore such capabilities in detail.

Damage Type

The methods proposed in this work are not based on any assumption regarding the type of the damage. The objective is to develop methods that can detect occurrence of structural abnormality in the surface of the pipe, regardless of their type/shape. Below is the list of structural abnormalities/defects used in this work to investigate different aspects of the proposed methods, and to validate their detection performance:

- (1) Two different sizes of mass scatterer to simulate structural abnormality: A light aluminum bar with 1.2 *cm* height and 1.2 *cm* diameter, and a heavier aluminum bar with 7.6 *cm* height and 5.08 *cm* diameter.

(2) Actual damage: A small cut (10% of the pipe's thickness) on the surface of the pipe to simulate crack, and a small mass loss (8% of the pipe's thickness). More details on the morphology and orientation of these cuts will be provided in later sections.

EOCs

For the laboratory experiments in this work, temperature is intentionally controlled to produce variations between 24°C and 39°C. In the laboratory experiments, temperature is the only varying EOC. However, since the objective of the proposed methods is to address the sensitivity of guided-waves to realistic EOC variation scenarios, these methods are validated using field data captured from an operating hot-water piping system, in which the list of varying EOCs is as follows:

- Coupling flowing water inside the pipe
- Temperature
- Fluid flow rate
- Inner pressure
- Mechanical noise

Pipe Material and Size

The physics behind the proposed methods is valid for different sizes of pipe, as long as the pipe's geometry supports propagation of guided-waves. In this work, three different sizes of pipes are used for experimental investigations and validations: (1) a Schedule-40 aluminum pipe segment, with 101.6 *mm* outer diameter and 5.8 *mm* thickness, (2) a Schedule-40 steel pipe with 33.5 *mm*

outer diameter and 3.3 *mm* thickness, and (3) a Schedule-40 steel pipe with 254 *mm* inner diameter and 9.27 *mm* wall thickness. In this work, only aluminum and steel pipes have been considered for experimental analysis. The extensibility of the results to other materials is not studied.

Chapter 3

Experimental Setups

3.1 Aluminum Pipe with Mass Scatterer

To control the temperature variations, I designed and built a setup in laboratory, which consists of a $1.5m \times 1.5m$ box made of insulation foam, with 50.8 mm thickness, R-value of $10\text{ (ft}^2\text{°Fhr/Btu)}$, and maximum operation temperature of 74°C . Figure 3.1 shows this setup. Interior temperature of the box is maintained within $\pm 0.5^\circ\text{C}$ of the specified setpoint, using a thermostatically controlled electric space heater. One of the pipe segments that I used for the experiments throughout this dissertation is a 1.2 m $3\frac{1}{2}$ Schedule-40 6061 T6 aluminum pipe, with 101.6 mm outer diameter and 5.8 mm thickness. Pitch-catch records are obtained using two Lead Zirconate Titanate (PZT) transducers that are coupled to the outer surface of the pipe. To record the interior temperature of the box, I located three HOBO H08-001-02 temperature data loggers at three points throughout the 1.2 m length of the pipe. Temperature readings show a uniform distribution of temperature along the pipe segment.

In order to excite guided-waves that are close to lamb waves in plates, the following criteria should be satisfied (Lefebvre et al., 2002; Protopappas et al., 2006): $r \gg h$, $\lambda \gg h$, and $r \gg \lambda$, where r , h , and λ are pipe radius, pipe thickness and wavelength, respectively. Therefore, for this aluminum pipe segment used in the laboratory, the desirable range for excitation frequency is approximately between 125 KHz and 1000 KHz. In all the laboratory experiments in which this pipe segment is used, I transmitted 0.1 ms Gaussian excitation signal with central frequency of 250 KHz and recorded 10 ms of ultrasonic signals, with sampling frequency of 10 MHz, at the receiver.

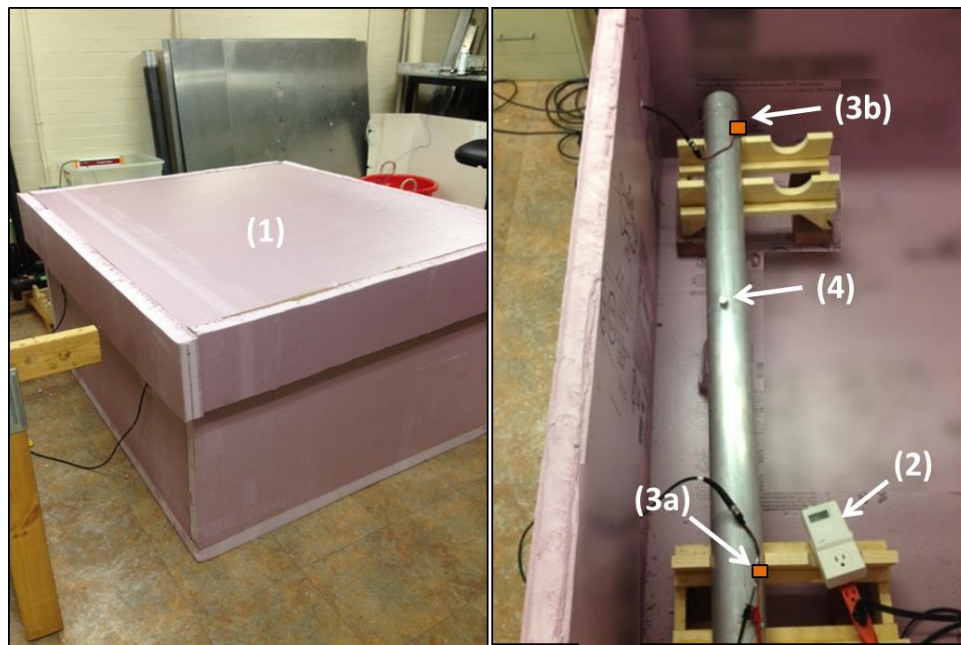


Figure 3.1: Laboratory setup for controlling temperature variation: 1) Thermally insulated box to contain the experimental pipe segments, 2) Thermostat, 3a & 3b) PZT transducer and receiver, respectively, 4) Grease-coupled mass scatterer to simulate a structural abnormality.

Structural abnormalities are simulated by masses of different sizes, grease-coupled to the outer surface of the pipe. Figure 3.2 shows the four different layouts for the aluminum pipe segment

that are used in the experiments throughout this work. In Figure 3.2a and 3.2b, transducers are located as far as 1.0m from each other. I refer to this transducer layout as layout #1. In Figure 3.2c and 3.2d on the other hand, the transducers are as close as one-third of the length of the pipe (0.4m). I refer to this transducer layout as layout #2. The perpendicular distance of the mass to the undamaged path A-B in Figure 3.2b and 3.2d is two times the distance in Figure 3.2a and 3.2c. I refer to the closer mass location (Figure 3.2a and 3.2c) as *Loc1*, and the further location (Figure 3.2b and 3.2d) as *Loc2*. For all the four layouts, data is collected using two different sizes of mass to reflect different sizes and characteristics of structural abnormalities. The small mass is a light aluminum bar with 1.2 cm height and 1.2 cm diameter, and the bigger mass is a heavier aluminum bar with 7.6 cm height and 5.08 cm diameter. For simplicity, throughout this work, these eight layouts will be referred to as “[1 or 2]-[Loc1 or Loc2]-[sml or big]”, where the first number indicates the transducer layout, the second symbol refers to the location of the mass, and the third symbol indicates the size of the mass.

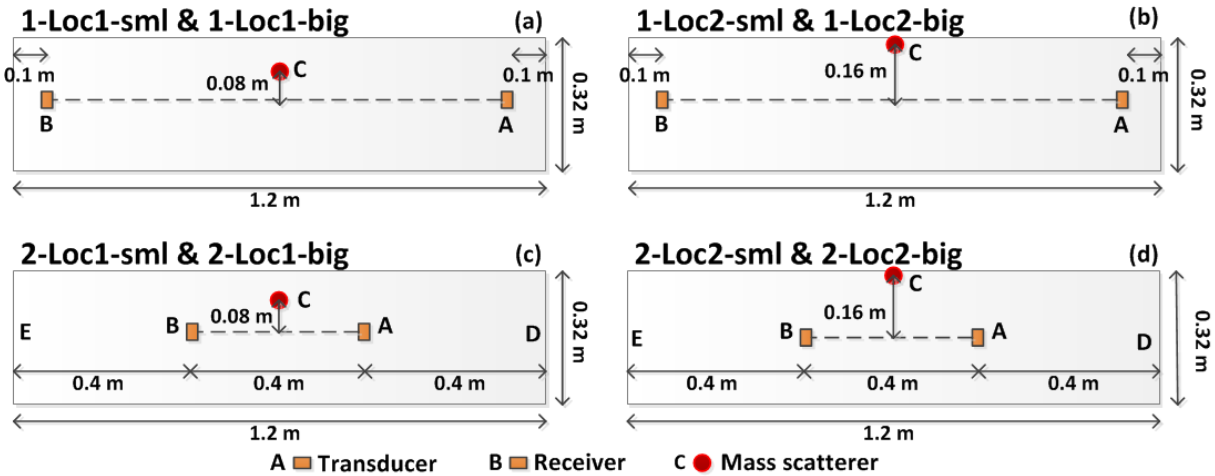


Figure 3.2: Unrolled view of the laboratory aluminum pipe segment, for different transducer-mass layout used throughout this work. Note: The three symbols in the layout names (1 vs. 2, Loc1 vs. Loc2, and big vs. sml) indicate the transducer setup, mass location, and mass size, respectively.

3.2 Steel Pipe with Crack and Mass Loss

This setup includes a Schedule-40 steel pipe (Figure 3.3) with 33.5 *mm* outer diameter and 3.3 *mm* thickness. Before introducing the crack, I obtained ultrasonic pitch-catch records from the intact pipe and from the pipe with a small mass scatterer placed in a non-symmetric location shown in Figure 3.3, at temperatures ranging from 24°C to 32°C. Later, I impose an oblique cut of 2.5 *cm* long, extended in both circumferential and longitudinal directions of the pipe (see Figure 3.3), at the middle of the length of the pipe, using a jewelers saw. The maximum thickness along the length of the crack is approximately 0.34 *mm* (10% of the pipe thickness), and the maximum depth is approximately 0.7 *mm* (20% of the pipe thickness). The second damage is a small mass loss (Figure 3.3) with 5 *mm* diameter and maximum depth of 0.25 *mm* (~8% of the pipe thickness). The location of the second damage differs from both the crack, and the mass scatterer location. Pitch-catch signals from the pipe with crack, and with both crack and mass loss are captured at $24^{\circ}\text{C} \leq T_{tst} \leq 32^{\circ}\text{C}$.

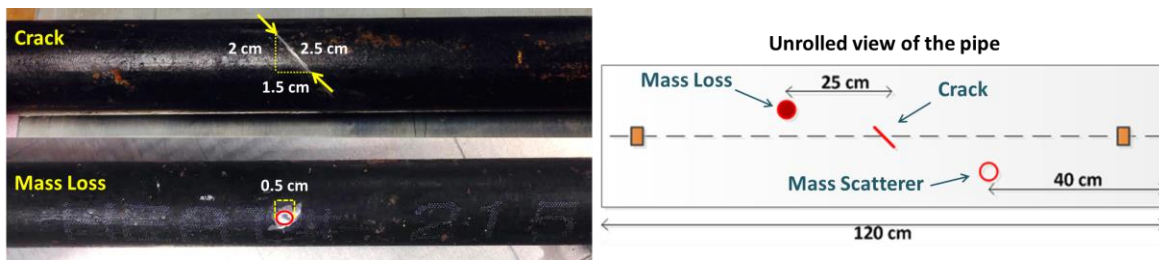


Figure 3.3: The steel pipe used to examine the application of the proposed methods for detection of multiple actual damages, namely a crack and mass loss.

3.3 Steel Pipe Operating Under Multiple Varying EOCs

For validation of the methods proposed in this work, the field data is obtained from a fully operational large-scale pressurized hot water piping system in the mechanical space of a campus building (Liu et al., 2012b). The characteristics of this testbed are completely different from the laboratory setups explained in the previous sections, which strengthen the generality of the validations.

The mechanical space is 706 m^2 and is mechanically and electrically noisy. This piping system is operating continuously and, unlike the laboratory pipes, is coupled with the flowing water with varying temperature, flow rate, and pressure. Due to the periodic pumping of hot water, the flow rate continuously varies between $45.5\text{ m}^3/\text{h}$ and $102\text{ m}^3/\text{h}$, and water temperature fluctuates from 38°C to 60°C . It is notable that the temperature variation for the laboratory experiments ranges from a minimum of 24°C to a maximum of 38°C . Therefore, the types/ranges of EOC variation are completely different from the laboratory data. The size and material properties of this pipeline are also different from the pipe segment used in the laboratory. This is also a Schedule-40 steel pipe with 254 mm inner diameter and 9.27 mm wall thickness, covered by fiberglass insulation.

Pairs of PZT transducers are permanently mounted on the pipe's exterior surface, 3 m or 6 m apart from each other. These distances are 2.5 to 15 times longer than the 1.2 m and 0.4 m ranges used for the laboratory aluminum and steel pipes. To simulate damage, a small aluminum bar (1.2 cm diameter and 1.2 cm height) is acoustically coupled to the surface of the pipe at $1/3$ of

the monitoring range. The pipe is excited with 0.1 *ms* broadband sinc signals with the frequency band of 100-300 kHz. This translates to frequency-thickness rate of 927-2781 kHz-*mm* as opposed to 1450 kHz-*mm* in laboratory experiments. The received 10 *ms* of ultrasonic signals are sampled with a sampling frequency of 10 MHz.

Chapter 4

A Supervised Approach for Damage Detection of Pipelines

4.1 Introduction

As discussed in Section 1.3, the challenges of guided-wave based damage detection are rooted in the complex nature of these waves and their sensitivity to EOC variations. The objective of this work is to simplify guided-waves, while retaining damage information. The idea is to extract damage-sensitive features from the simplified signals so that they are less sensitive to EOC variations.

A supervised method is proposed to extract a sparse subset of the guided-wave signals that contain optimal damage information for detection purposes. Ideally, if the arrivals scattered from damage are completely retrieved from the recorded signals, they could be used for damage detection. However, studies have shown that variation of EOCs affects scatter signals so that

damage information are suppressed by the EOC effects (Croxford et al., 2010; Eybpoosh et al., 2015, 2014b; Lu and Michaels, 2005). The fundamental assumption for development of the proposed method is that, by maximizing the detection performance while imposing sparsity constraint, in the extracted sparse signals, the effects of damage are more dominant than the EOC effects. If true, this damage-sensitive subset can be used for damage detection rather than the complete scatter signal. This assumption, as well as different aspects of the proposed method, is validated later in this chapter.

4.2 Motivation

Being reflected from any scatterers in the structure (*e.g.*, welding, geometric boundaries, damage), guided-waves in a medium travel through multiple paths. Depending on the wave velocities, travel path, boundary conditions, *etc.*, these reflections will arrive to the receiver at different ranges of time throughout the sampling period. Any point in time will include different portions of these arrivals, either individually or as superposition of multiple arrivals. However, it is important to note that, not all these arrivals contain significant damage information. Intuitively, an arrival that has only illuminated the undamaged section of the medium will not contain significant information regarding the damage (note: this is true only if the damage size is small enough that the physical properties of the medium such as rigidity and Young's modulus are not changed significantly).

For example, considering a pipe with small wall thickness to diameter ratio (Pierce and Kil, 1990), let Figure 4.1 be an unrolled view of the pipe, with a two-transducer pitch-catch setup,

where excitation happens at point A, and arrivals are recorded at point B. Let the small circle at point C be the damage on the pipe, and the two ends of the pipe symbolize any scatterer in the pipe, such as welding, boundaries, *etc.* Note that this figure is just a simple illustration and does not include all of the possible wave propagation paths.

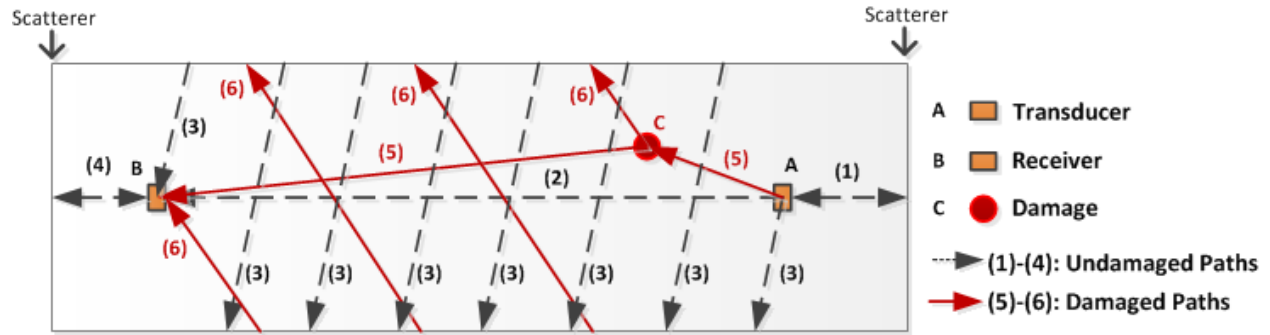


Figure 4.1. A schematic unrolled view of a pipe, illustrating example wave propagation paths passing through undamaged (dashed arrows) and damaged (solid arrows) sections of a pipe.

In this schematic depiction (Figure 4.1), arrivals that have travelled through paths 1 to 4 (*i.e.*, the paths that do not include point C) are not expected to contain significant information regarding damage. On the other hand, arrivals that have illuminated the damaged section of the pipe (*i.e.*, paths 5 and 6 that include point C) may contain damage information. Any record during sampling period that contains part of these arrivals would include information regarding damage. However, the question is that in what subset of such records this information suppresses the irrelevant information and leads to optimal damage detection under varying EOCs.

4.3 Method

The goal is to extract a subset of the guided-wave signals that contains enough information regarding existence of damage. The optimal solution will be the “simplest” subset, with minimum sensitivity to EOC variations, and maximal damage information for detection purposes. The main hypotheses of the proposed method are as follow:

Hypothesis #1: A sparse subset of the arrivals in a diffuse-field signal contains enough information for optimal damage detection.

Hypothesis #2: If hypothesis #1 is true, the extracted sparse subset is less sensitive to EOC variations and more sensitive to damage than the complete signal.

These hypotheses will be addressed through research questions #1 and #2 presented in Section 4.3.1, and will be verified through a variety of experiments reported later in this chapter. A supervised approach is proposed to find a discriminant vector in the time space of the signals so that the signals from damaged and undamaged pipes have different projections onto this vector. In other words, the vector obtained by this method (will be referred as sparse discriminant (SD) method) is located in a direction that the arrivals with significant damage information have distinguishable projections onto the vector.

4.3.1 Research Questions

Based on the discussions in the motivation section (Section 4.2), and the hypotheses given in Section 4.3, following is the list of research questions that need to be answered:

Question #1: How can a sparse representation of a diffuse-field guided-wave signal in a pipe (with specifications summarized in Section 2.2) be extracted while retaining enough damage information for detection purposes?

Question #2: How sensitive are the extracted subset and any damage-sensitive feature used in Question #1 to the variations of EOCs specified in Section 2.2 of this manuscript?

4.3.2 Overview of the Sparse Discriminant (SD) Method

Proposed SHM Framework based on the SD Method

Figure 4.2 illustrates the application of the proposed SD method for continuous, online damage detection of a pipeline. The process consists of an initial training stage, and the continuous monitoring stage. Different components of these stages are introduced in this section, and experimentally examined in the next sections of this chapter.

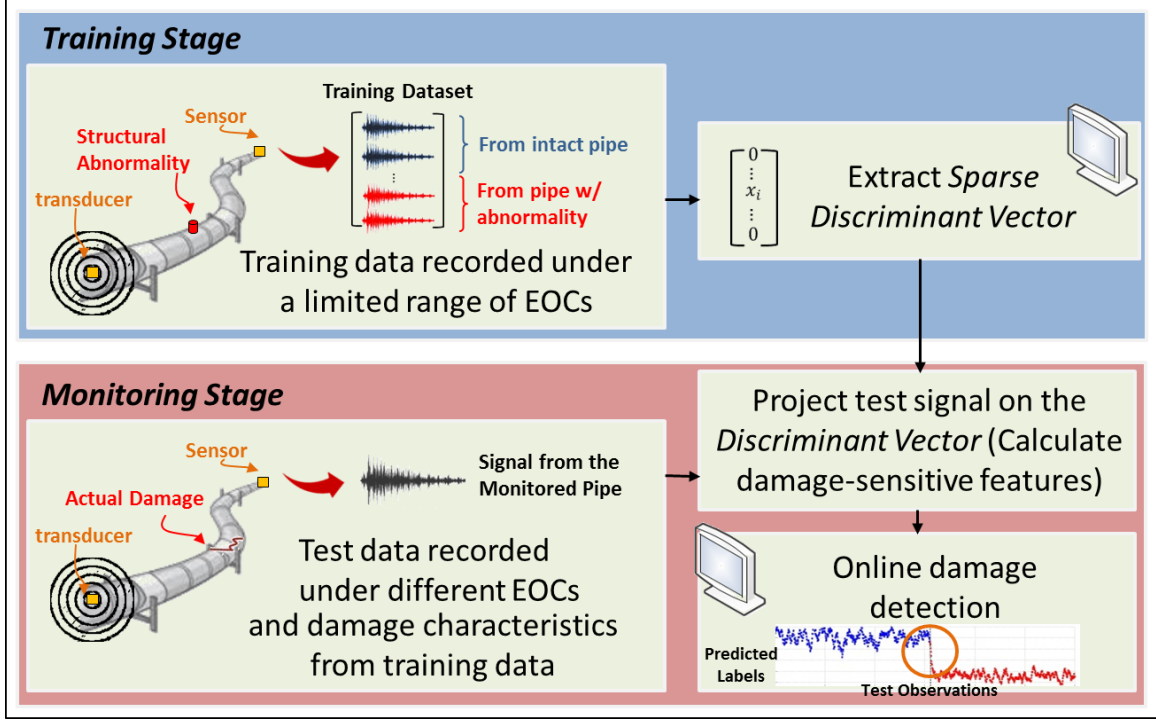


Figure 4.2: Application of the proposed SD method for continuous monitoring of pipelines.

Training Stage

As discussed in Section 1.3, when guided-waves are interfered by damage, their propagation characteristics may change in a number of ways, such as mode conversion, change in the phase/group velocities, multi-path reflections, energy dissipation, etc. (Belanger and Cawley, 2009; Dehghan-Niri and Salamone, 2014; Nagy et al., 2014). At this point it, it is notable that the proposed method do not make any assumption regarding the nature of such interactions, but rather its goal is to extract a subset of the signal with optimal damage information, regardless of the physical phenomenon leading to such a subset.

As shown in equation 4.1, the coefficients of a discriminant vector ($\vec{X}^{n \times 1}$) are trained so that the projections of training signals ($Z^{m_{tr} \times n}$) on $\vec{X}^{n \times 1}$ are good predictors of the state of the pipe (*i.e.*,

$Y_j = 1$ for intact pipe and $Y_j = -1$ for pipe with structural abnormality, where Y_j is the label of the j th observation). Let m_{tr} be the number of signals in the training dataset, including signals from intact and damaged pipe, n be the length of the time-trace of the signals, and Y be the vector of state labels.

$$\operatorname{argmin}_x \{ \|Z^{m_{tr} \times n} \vec{X}^{n \times 1} - \vec{Y}^{m_{tr} \times 1}\|_2 \} \quad (4.1)$$

In equation 4.1, if the sample points in training signals are zero-meaned and normalized by standard deviation (*i.e.*, standardize Z in column), the magnitude of the coefficients in \vec{X} will reflect the significance of each of the n sample points in defining the state of the pipe.

At this point, it is notable that *linear discriminant analysis* (LDA) is another method used in pattern recognition to find such linear subspaces that separate different classes. However, as discussed in one of the authors' work (Eyμποosh et al., 2014b), LDA fails to find such a discriminant subspace for guided-waves when using the original time-trace of the signals. It is mainly because LDA assumes that the variables are normally distributed. However, as I will further discuss in Section 4.5.4, guided-wave signals do not satisfy this assumption.

Although the method proposed in equation 4.1 does not assume any particular distribution for the n sample points, its performance can still be affected by the curse of dimensionality. If the number of training observations (m_{tr} , slow-time recording duration) is smaller than the number of predictor variables, n (thousands of sample points in a signal, *i.e.*, fast-time recording duration), the extracted coefficients in the \vec{X} vector may not be statistically significant. Satisfying

the $m_{tr} > n$ condition is neither practical, nor favorable. Therefore, the dimensionality of the problem needs to be reduced.

Common methods for dimensionality reduction in time-domain, such as down-sampling, filtering, *etc.*, would be vulnerable to removing useful information, or would require prior knowledge about the propagating wave modes, wave reflection scenarios, damage properties, *etc.* The method proposed in this study assumes that a *sparse* subset of a guided-wave signal will contain sufficient damage information for optimal damage detection. This assumption is in line with the discussions provided earlier in Section 4.2 (Figure 4.1), and will be verified further in the validation section of this chapter (*i.e.*, Section 4.5).

The suggestion is to penalize the magnitude of the coefficients that correspond to the variables with less contribution in predicting the class labels \vec{Y} (using a regularization scalar ξ in equation 4.2). Ideally, this can be done by penalizing the ℓ_0 norm of the \vec{X} vector ($\|\vec{X}\|_0$). That is, forcing the optimization algorithm to assign zero coefficients to the variables with insignificant contribution in reconstructing \vec{Y} , and non-zero otherwise. However, finding the solution for such a problem is NP-hard (Non-deterministic Polynomial-time hard) (Huang and Aviyente, 2006). An approximate solution to this problem can be found by penalizing ℓ_1 norm of the \vec{X} vector instead of ℓ_0 norm (Donoho and Huo, 2001). This forces a sparse solution for the \vec{X} vector, so that, variables with smaller contribution in defining the state of the pipe will be assigned *close-to-zero* coefficients.

$$\operatorname{argmin}_x \{\|Z^{m_{tr} \times n} X^{n \times 1} - Y^{m_{tr} \times 1}\|_2 + \xi \|X^{n \times 1}\|_1\} \quad (4.2)$$

Finding a *sparse* discriminant vector (\vec{X}) in the time space of the signals addresses the complexity challenges of guided-waves discussed in Section 1.3 (through the second part of equation 4.2), while retaining damage information for detection purpose (through the first part of equation 4.2). The question remained to be answered is whether the effects of damage in the extracted sparse subset of the signals suppress the EOC effects under different scenarios.

It is notable that equation 4.2 is based on the Lasso optimization principle which is different than ordinary regularization methods such as Tikhonov regularization, since Lasso minimizes ℓ_1 norm instead of Euclidean norm, and in this way, the sparsity is imposed to the solution. For the experiments reported in this chapter, the Matlab convex optimization package (*i.e.*, *cvx* by Grant and Boyd (2008)) is used to solve the Lasso optimization problem formulated in equation 4.2. The regularization scalar ξ can be selected so that the sparsity is maximized while training error is minimized.

Monitoring Stage

The proposed SD method can be used to extract damage-sensitive features for damage detection during the monitoring stage depicted in Figure 4.2. Projecting a test signal \vec{d}_j ($\vec{d}_j \in R^n, j = 1, \dots, m_{tst}$) on the trained \vec{X} vector will result in a scalar (equation 4.3) representing the predicted class label for the signal (ideally, $\hat{Y}_j = 1$ for intact pipe and $\hat{Y}_j = -1$ for pipes with structural abnormalities).

$$\hat{Y}^{m_{tst} \times 1} = D^{m_{tst} \times n} \vec{X}^{n \times 1} \quad (4.3)$$

4.3.3 Evaluation Criteria for the SD Method

Detection Performance

EOC variations, along with other parameters that may vary between training and monitoring stages (*i.e.*, damage characteristics and location), may cause the values of \hat{Y} s deviate from exact 1 and -1. The first objective to be evaluated is the extent to which the extracted sparse discriminant vector, and hence the damage-sensitive features (\hat{Y} s in equation 4.3), retain damage information.

In order to examine the performance of these features in discriminating damaged and undamaged pipes, three metrics are used in this study:

- (a) Detection accuracy: the ratio of the number of correctly labeled damaged and undamaged observations to the total number of test observations.
- (b) False negative rate (FNR): the ratio of the number of incorrectly labeled damaged observations to the total number of damaged observations.
- (c) False positive rate (FPR): the ratio of the number of incorrectly labeled undamaged observations to the total number of undamaged observations.

At this point, it is worth emphasizing that the SD method is not a detection algorithm, but a feature extraction approach for damage detection. Therefore, a number of classification methods can be used to evaluate discriminatory power of the features. A simple method is to cluster the predicted labels. In this work, a 2-class k-means clustering is used for this purpose. K-means is a

simple unsupervised partitioning method in which each observation is assigned to one of the clusters. K-means clustering is used since it is an unsupervised method that do not need training data, therefore, its performance will be independent from training parameters. After clustering different test observations, some heuristic process is applied to calculate the three aforementioned detection metrics. If more than 50% of the observations in a cluster are from the same class (e.g., intact), the cluster is considered to be representative of that particular class. Then, accuracy, FPR, and FNR can be calculated. It is also notable that K-means algorithm can be very sensitive to the initial values of cluster centroids. The results reported in this chapter are the average of different folds of cross-validations. Also, the initial centroid values are not optimized, which suggests that the reported findings can be even further improved if more sophisticated classifiers are applied.

Sparsity

In order to quantify the simplicity of the extracted sparse signals, for different scenarios, *i.e.*, to verify the sparsity assumption of the SD method, I introduce a metric, Sparsity-ratio (Sr), which is the ratio of the number of zero coefficients to the total length of the signal n (equation 4.4). It is notable that, using the ℓ_1 norm of the \vec{X} vector in equation 4.2, the values of the coefficients may never be absolute zero, but rather be very small. To handle this, any coefficient whose magnitude is smaller than the largest coefficient (*i.e.*, $x_{noise}^{max}(\xi)$) assigned to the initial part of the signal (*i.e.*, before the first arrivals) is considered to be *zero*. This is because this part of the signal is expected to have no contribution in defining the state of the pipe. The $Sr(\xi)$ given in equation 4.4 is the sparsity ratio corresponding to the regularization scalar ξ .

$$Sr(\xi) = \frac{|\{i \in \{1, 2, \dots, n\} : |x_i| \leq |x_{noise}^{max}(\xi)|\}|}{n} \times 100 \quad (4.4)$$

Sensitivity to EOCs

The second objective to be evaluated is low sensitivity of the SD method to EOC variations. To evaluate this, I examine: (a) detection performance, and sparsity ratio (Sr), as the SD method is trained at different EOCs, and (b) detection performance as the EOCs between test and training data vary. It is notable that for the laboratory experiments used to evaluate the SD method, temperature is the only varying environmental factor. However, the performance of the SD method is later validated using field data recorded from a pipeline operating under dynamic conditions with wide range of varying temperature, fluid flow rate, and inner pressure, among others.

4.4 Experimental Investigation of the SD Method

This section provides a general proof of concept for the proposed SD method. The basic assumptions of the SD method, research questions, and different aspects of the framework proposed in Figure 4.2 are studied using the experimental data captured from the aluminum pipe discussed in Section 3.1.

4.4.1 Temperature Variation

Detection Performance of Damage-sensitive Features

As shown in equation 4.3, the trained coefficients can be used to predict class labels of new observations. In this section, I examine whether, for different temperature variation scenarios, the class labels predicted by SD method can be used as features to detect the structural abnormalities.

In order to investigate the effect of temperature difference (ΔT) between training and test datasets, 5,500 observations from intact and 5,500 observations from the pipe with abnormality are measured, at temperatures ranging from 24°C to 32°C, from *I-Loc1-sml* pipe layout (Figure 3.2a). A total of 55 datasets, each consisting of 100 intact and 100 damaged signals are created. The coefficients are trained and tested with each of these datasets at a time.

As mentioned before, the focus of this study is not to propose a particular classifier for damage detection, but to propose an approach for extracting damage-sensitive features for damage detection. The discriminatory power of these features (\hat{Y} s) can be tested using various classification methods. In this work, a simple K-means clustering method found to be satisfactory to separate the predicted test labels.

A total of 6,050 training/test scenarios are examined (55 training datasets, each tested by 55 test datasets, through 2-fold cross validation CV). The ij th pixel in Figure 4.3a is the average accuracy of 2-fold CV for the i th testing dataset with temperature T_i and the j th training dataset

with temperature T_j . As can be seen in this figure, for the majority of the scenarios, two classes are perfectly separated (detection accuracy is 100%). However, expectedly, detection performance drops slightly as ΔT between training and testing increases (the lower right and upper left corners of the figure). That is, the extracted sparse vectors become less representative of testing data as ΔT increases. The FPRs and FNRs are below 10% for 97% of the scenarios.

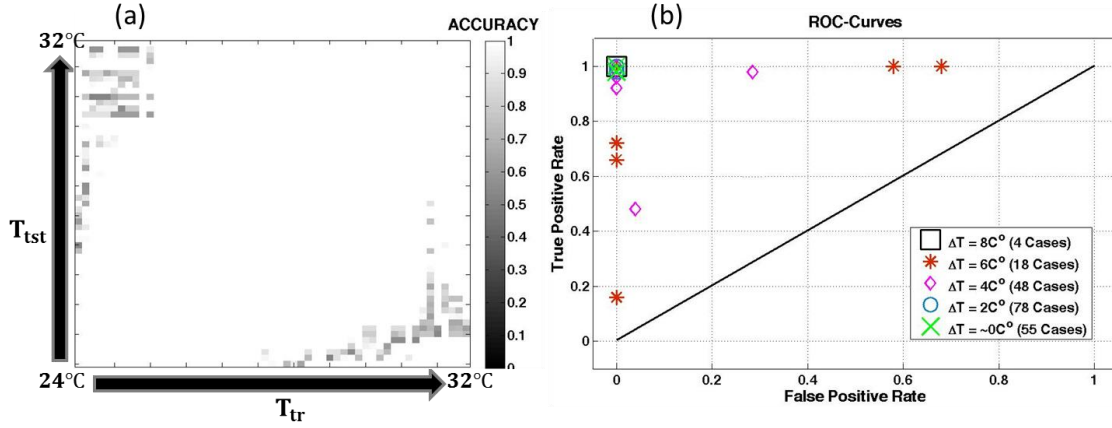


Figure 4.3: (a) Average detection accuracy of 2-fold CV, for different training/test temperature combinations ($24^\circ\text{C} \leq T \leq 38^\circ\text{C}$), using the class labels predicted by SD method (\hat{Y} s) as the only damage-sensitive feature for clustering. (b) ROC-curves for training/test scenarios with $0^\circ\text{C} \leq \Delta T \leq 8^\circ\text{C}$. T_{tr} : Training temperature, T_{tst} : Test temperature.

Figure 4.3b shows the receiver operating characteristic (ROC) curves for different cases in which training and test temperatures vary between 0.0°C and 8.0°C . For almost all these cases, these curves are far from the 45-degree diagonal of the ROC curve, suggesting high sensitivity of the test (TPR) even with the drop in specificity (1-FPR). These results imply that, not only the predicted labels (\hat{Y} s) are significantly damage-sensitive, but also, the separation between \hat{Y} s of two classes has low sensitivity to ΔT (*i.e.*, two classes can still be separated for large ranges of

ΔT s). Table 4.1 summarizes the average detection statistics for 2-fold CV of all 3,025 training/test combinations with different ΔT scenarios.

Table 4.1: Average statistics of 2-fold CV, reflecting the detection performance of class labels predicted by SD method (\hat{Y} s), for a total of 3,025 training/test scenarios given in Figure 4.3.

Average Detection Accuracy	Average FPR	Average FNR
99.0%	1.0%	1.4%

It is useful to have a closer look at the temperature effects on the predicted class labels. The distinction between intact and damaged observations is related to the distance between \hat{Y} s in two classes. For each training/test combination, Figure 4.4 shows the change in the distance between average predicted labels for intact and damaged test observations, as a ratio of the standard deviation of the predicted labels for intact observations ($dist = |\bar{\hat{Y}}_{int} - \bar{\hat{Y}}_{dmg}| / \sigma_{\hat{Y}_{int}}$). The brighter colors indicate larger distance between the labels, and hence, clearer distinction between the two classes. These distances can get as large as 350 times the σ_{int} , and for the worst training/test scenario, as low as two times the σ_{int} . Expectedly, the distinction between predicted labels degrades as the ΔT increases (*i.e.*, moving further from diagonal of the matrix in Figure 4.4). The wider dark region in the lower triangle of the figure, compared to upper side, implies that the degradation in the distinction between the class labels may also be sensitive to training temperature in addition to ΔT . Large *dist.* values for the majority of the scenarios implies that the class labels predicted by SD method can perform reasonably even at large ΔT s (up to 8°C in this experiment).

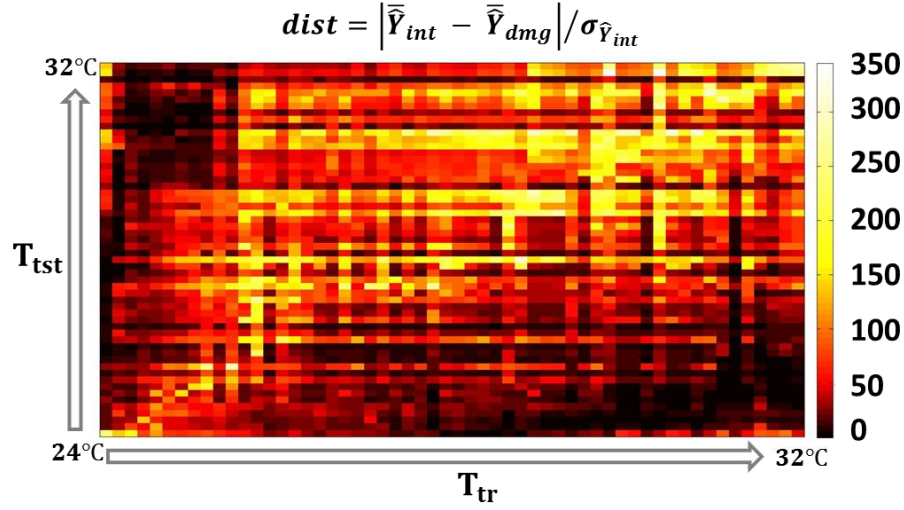


Figure 4.4: Variation of the distance between predicted labels for intact and damaged test observations at different training/test temperature scenarios. T_{tr} : Training temperature, T_{tst} : Test temperature, \bar{Y}_{int} : Average predicted labels for intact test observations, \bar{Y}_{dmg} : Average predicted labels for damaged test observations, $\sigma_{\bar{Y}_{int}}$: Standard deviation of the predicted labels for intact test observations.

Online Damage Detection

Low temperature sensitivity of the predicted class labels makes them attractive for online damage detection of pipelines. The state of a pipe operating under varying temperatures can be predicted as the new observations are streamed (equation 4.3). This implementation of the SD method is illustrated with an example in Figure 4.5. In this example, training data include 50 intact and 50 damaged observations, all measured at 26°C, from the *I-Loc1-sml* layout (Figure 3.2a). Test dataset includes 550 observations from intact pipe, and 550 observations from the *I-Loc1-sml* pipe, in which the abnormality is simulated by introducing a small mass. As can be seen in Figure 4.5a, temperatures of the test observations can differ significantly from the training temperature ($-4^{\circ}\text{C} \leq \Delta T \leq +5^{\circ}\text{C}$). Figure 4.5b shows the correlation between each test

signal and the average training intact signal. No distinguishable pattern in the correlations of intact versus damaged test signals is observed.

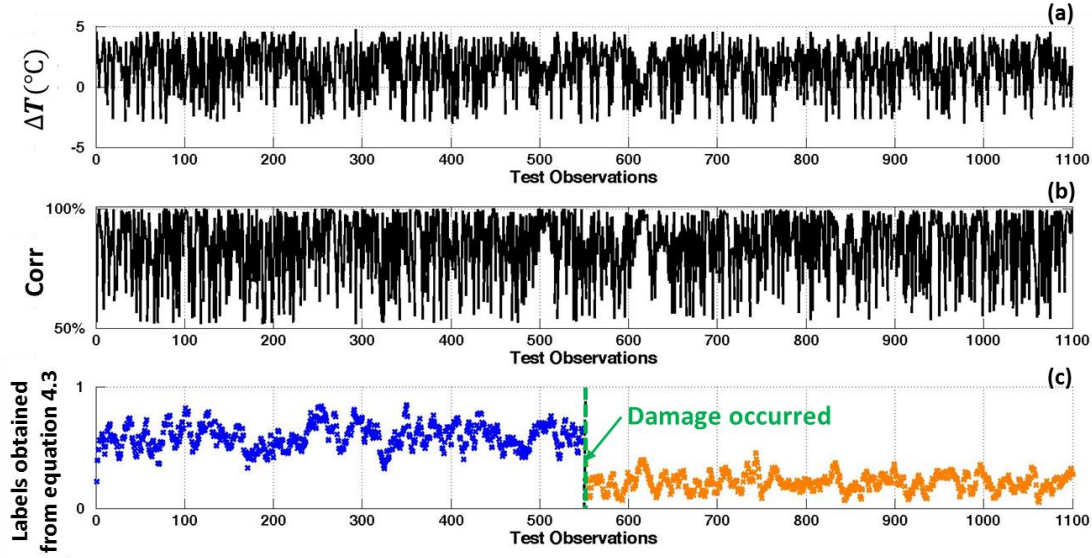


Figure 4.5: Illustration of online damage detection with varying temperature, using class labels predicted by SD method. (a) Temperature difference between the training dataset and the monitored observations, (b) Correlation between the monitored observations and average of training intact signals, (c) Predicted class labels for the monitored observations using equation 4.3.

The online detection discussed here could be performed by updating the training dataset as the test observations are streamed. However, in this study, in order to investigate the performance of the SD method when the EOCs are different for test and training data, the training dataset, and thus the calculated \vec{X} vector, are not updated. In other words, the online monitoring implementation reported in this work is not adaptive.

To quantify the detection performance of the predicted labels (\hat{Y} s) for online monitoring, a simple detection algorithm is used. First, predicted labels of the test observations are averaged

with a window of 20 records, which results in the values given in Figure 4.5c. It is notable that the observations are recorded in 1-minute intervals. When the distance between the label of the j th observation (\hat{Y}_j) and the average of the labels in the window before the j th observation (\hat{Y}_{j-20} to \hat{Y}_{j-1}) is larger than ten times the standard deviation of the labels in the window, occurrence of damage is detected ($\Delta\hat{Y} > 10\sigma$). This approach can be further improved by using the results given in Figure 4.4. That is, adjusting the standard deviation threshold based on the ΔT between the monitored signal and training data.

Table 4.2: Detection performance of \hat{Y} s for online monitoring of pipes, under varying temperatures. Note: *1-Loc1-sml* layout is used. Test data includes 610 intact and 610 damaged observations, recorded in 1-minute intervals.

Training Temperature	Range of ΔT	Delay in Detection (No. of observations before detection)	Detection Accuracy
26°C	$-2^\circ\text{C} \leq \Delta T \leq 13^\circ\text{C}$	9	99.2%
27°C	$-3^\circ\text{C} \leq \Delta T \leq 12^\circ\text{C}$	2	99.8%
28°C	$-4^\circ\text{C} \leq \Delta T \leq 11^\circ\text{C}$	6	99.5%
29°C	$-5^\circ\text{C} \leq \Delta T \leq 10^\circ\text{C}$	2	99.8%
30°C	$-6^\circ\text{C} \leq \Delta T \leq 9^\circ\text{C}$	6	99.5%
31°C	$-7^\circ\text{C} \leq \Delta T \leq 8^\circ\text{C}$	3	99.7%
32°C	$-8^\circ\text{C} \leq \Delta T \leq 7^\circ\text{C}$	6	99.5%
Average		4.8	99.5%

Table 4.2 summarizes the online monitoring results, when SD method is trained at different temperatures, ranging from 26°C to 32°C. The test dataset includes a total of 1,220 intact and damaged signals at wide range of ΔT s. In all these cases, occurrence of damage is detected, with an accuracy of above 99%. It is notable that the detection statistics of the simple algorithm

explained above ($\Delta\hat{Y} > 10\sigma$) may vary depending on the selected window size and/or the σ threshold. While the results given in Table 4.2 prove the concept, the online detection performance can be improved even further if more sophisticated approaches are used for detection of the divergence in values of \hat{Y} s.

Evaluation of the Sparsity Assumption

Generally, no significant correlation could be observed between sparsity ratio (equation 4.4) and the temperature of the training dataset in my experiments. Table 4.3 summarizes the statistics for the sparsity ratios of training datasets recorded from *I-LocI-sml* pipe layout (Figure 3.2a) at $24^{\circ}\text{C} \leq T_{tr} \leq 38^{\circ}\text{C}$. The results show that, regardless of the temperature of the training dataset, the solution of equation 4.2 is sparse, which, as shown earlier in this section, leads to high detection performances.

Table 4.3: Sparsity ratio (Sr., equation 4.4) statistics as a result of training the SD algorithm with datasets of different temperatures.

Range of Temperature	Average Sr.	Minimum Sr.	Maximum Sr.	Standard Deviation
$24^{\circ}\text{C} \leq T_{tr} \leq 38^{\circ}\text{C}$	90.0%	60.0%	99.7%	10.0%

4.4.2 Damage Size and Temperature Variation

Considering a particular type of damage positioned in a particular location, the arrival time of the waves reflected from damage boundaries is determined by the distance of the boundaries to the transducers. The size and shape of the damage define the location of damage boundaries with respect to the transducers. For example, consider a case where the structural abnormality used for

training is so large that the front and back edge reflections can be separated in time. Intuitively, the trained coefficients in this case will not be a good representative for a test observation in which the damage is so small that the front and back edge reflections arrive around the same time, and are mostly overlapped.

Wang et al. (2010) have used the distinction between the reflections from damage edges to characterize the extent of the damage. Their approach, however, require damage size to be large enough so that the front and end reflections are separable. For example, they show that, in order for the two damage edge reflections to be separable, minimum longitudinal extent of the damage should be around 86 *mm* and 170 *mm*, for 175 KHz and 200 KHz excitations, respectively. These values correspond to 2.5 and 5 times the diameter, and 4% and 8% of the length of the pipe segment used in Wang et al. (2010). Obviously, real-world pipeline monitoring applications require detection of defects much smaller than these ranges. This implies that, in realistic detection scenarios, the reflections from damage edges are expected to arrive around the same time and be highly overlapped. Therefore, the time locations of the coefficients trained with different sizes of damage will not vary significantly (assuming that the damages are of the same type and are located at the same point in pipe). In light of this, the hypothesis being tested in this section is as follows:

Hypothesis #3: For practical sizes of damage, the difference in damage size between training and test data will not adversely affect the detection performance of \hat{Y} s predicted by the SD method.

Detection Performance of Damage-sensitive Features

To test hypothesis #3 given above, the first set of experiments is to investigate detection performance of \hat{Y} s when the size of structural abnormality in training and test data is different. Two different sizes of mass are used to simulate the variation in the size of the structural abnormalities. As explained in Section 3.1, the height and diameter of the bigger mass are about six and four times bigger than the small mass, respectively. If hypothesis #3 is verified for such a large difference, the findings can be safely expanded to smaller variations in the size of damages with similar characteristics as these experiments.

A total of 25 training datasets ($25^{\circ}\text{C} \leq T_{tr} \leq 33^{\circ}\text{C}$) are created. Each training dataset consists of 100 observations from intact pipe and 100 observations from pipe with small mass (*1-Loc1-sml*, Figure 3.2a). Similarly, a total of 25 test datasets ($25^{\circ}\text{C} \leq T_{tst} \leq 33^{\circ}\text{C}$) are created. Each test dataset consists of 100 observations from intact pipe and 100 observations from pipe with big mass (*1-Loc1-big*, Figure 3.2a). The coefficients are found for all 25 training datasets, and tested with all 25 test datasets, using K-means with 2-fold CV. This translates to a total of 1,250 training/test scenarios. Table 4.4 summarizes the average detection statistics for all scenarios. It can be seen that the discriminatory power of \hat{Y} s is not affected by variation of damage size, or temperature, between training and test datasets.

Table 4.4: Average detection statistics for a total of 1,250 training/test scenarios with different damage sizes and temperatures. Note: $25^{\circ}\text{C} \leq T_{tr} \leq 33^{\circ}\text{C}$. Pipe layout *1-Loc1-sml* is used for training and layout *1-Loc1-big* is used for testing.

Average Detection Accuracy	Average FNR	Average FPR
99.0%	1.0%	0.9%

Online Damage Detection

Table 4.5 summarizes the performance of \hat{Y} s for online monitoring. The coefficients are found using seven different training datasets, at temperatures ranging from 26°C to 32°C (one training dataset for each temperature), using small mass (*1-Loc1-sml*). Test dataset includes 610 signals from intact pipe, 610 signals from pipe with small damage (*1-Loc1-sml*, Figure 3.2a), and 500 signals from pipe with big damage (*1-Loc1-big*, Figure 3.2a). The temperatures of the test observations vary randomly between 24°C and 39°C. As can be inferred from Table 4.5, although the coefficients are calculated using only the pipe with small damage, the occurrence of both small and big damages can be detected, for a wide range of ΔT s. This also suggests the potential of the SD method for detecting changes in the severity of damage.

Table 4.5: Detection performance of \hat{Y} s for online monitoring of pipes, when sizes of damage, as well as the temperatures of training and test datasets are different. Note: *1-Loc1-sml* pipe layout is used for training, and both *1-Loc1-sml* and *1-Loc1-big* layouts are used for testing. Observations are recorded in 1-minute intervals.

Training Temperature	Range of ΔT	Delay in Detection of Small Mass (No. of observations)	Delay in Detection of Big Mass (No. of observations)
26°C	$-2^{\circ}\text{C} \leq \Delta T \leq 13^{\circ}\text{C}$	10	6
27°C	$-3^{\circ}\text{C} \leq \Delta T \leq 12^{\circ}\text{C}$	3	6
28°C	$-4^{\circ}\text{C} \leq \Delta T \leq 11^{\circ}\text{C}$	6	6
29°C	$-5^{\circ}\text{C} \leq \Delta T \leq 10^{\circ}\text{C}$	3	7
30°C	$-6^{\circ}\text{C} \leq \Delta T \leq 9^{\circ}\text{C}$	5	3
31°C	$-7^{\circ}\text{C} \leq \Delta T \leq 8^{\circ}\text{C}$	5	4
32°C	$-8^{\circ}\text{C} \leq \Delta T \leq 7^{\circ}\text{C}$	14	0
Average		6.5	5.3

Evaluation of the Sparsity Assumption

Similar to temperature variations, the size of the structural abnormality found to be weakly correlated to the sparsity ratios. Table 4.6 summarizes the statistics for sparsity ratios obtained through training the SD algorithm with small and big damages, using *2-Loc2-sml* and *2-Loc2-big* pipe layouts (Figure 3.2d), respectively. As shown in this table, the solution of equation 4.2 is sparse for both sizes of damage.

Table 4.6: Sparsity ratio (Sr, equation 4.4) statistics as a result of training SD algorithm with signals from a pipe introduced to two different sizes of abnormality, at different temperatures ($24^{\circ}\text{C} \leq T_{tr} \leq 38^{\circ}\text{C}$).

Damage Size	Average Sr.	Minimum Sr.	Maximum Sr.	Standard Deviation
Small	91.0%	63.0%	99.8%	11.0%
Big	80.0%	60.0%	99.0%	12.0%

4.4.3 Damage Location and Temperature Variation

Difference in the location of damage in the pipe between training and test data is the third factor that may affect the performance of the SD method. Considering a particular type and size of damage, the position of the damage, with respect to the transducers, defines the arrival times of the waves reflected from damage. If damage locations in training and test data are so different that the arrival times from damage differ significantly, then the trained coefficients may not be a good representative of the subset of the test signals with significant damage information. In this section, we examine the impacts of damage location on detection performance of the SD method.

Figure 4.5a shows an unrolled view of a pipe with small wall thickness-diameter ratio (Pierce and Kil, 1990). Let C be the damage in the pipe. Here, I assume that the damage is located somewhere between the two transducers A and B, and both transducers are at the same elevation of the pipe. L_u is the monitoring range. The solid red arrows show example helical paths from transducer to damage, and from damage to the receiver.

A helical path of order n between two transducers located in horizontal distance of l and vertical distance of z can be obtained through equation 4.5 (Dehghan-Niri and Salamone, 2014; Nagy et al., 2014):

$$l_n = \sqrt{l^2 + (z + 2\pi nr)^2} \quad (4.5)$$

The length of any damaged path in Figure 4.5a is the sum of the path from transducer A to the damage C (L_d^s), and the helical paths of different orders (n) from the damage to the receiver B ($L_d^{h,n}$). To calculate the helical paths $L_d^{h,n}$, damage can be considered as a virtual transducer, with $l = L_u - y$. Therefore, damaged paths for the general case shown in Figure 4.5a are obtained as follows:

$$L_d = L_d^s + L_d^{h,n} = \sqrt{(\pi r - x)^2 + y^2} + \sqrt{(L_u - y)^2 + (z + 2\pi nr)^2} \quad (4.6)$$

In any damage scenario, the shortest damaged path is of particular importance, because it defines the time of the first arrivals. For any damage location, the shortest helical path is $L_d^{h,0}$, that is when $n = 0$, and damage is located in the same elevation as the receiver (*i.e.*, $z = 0$). Therefore, in Figure 4.5a, the minimum helical path would happen when the damage is located somewhere

on the AB path. In this case, the L_d^s is also minimal. For example, if $x = \pi r, y = L_u/2$ as shown in Figure 4.5b, referring to equation 4.6, minimum damaged path is as follows:

$$L_d^{min} = L_u \quad (4.7)$$

The maximum L_d can happen when damage is located in one of the pipe edges (i.e., $x \in \{0, \pi r\}, y \in \{0, L_u\}, z = 2r$). An example position is shown in Figure 4.5c. Based on equation 4.6, possible L_d^{max} values are as follow:

$$L_d^{max} = \begin{cases} \pi r + \sqrt{L_u^2 + 4r^2(1 + \pi n)^2}, & x = 0, y = 0 \\ \sqrt{\pi^2 r^2 + L_u^2} + 2r(1 + \pi n), & x = 0, y = L_u \\ \sqrt{L_u^2 + 4r^2(1 + \pi n)^2}, & x = \pi r, y = 0 \\ L_u + 2r(1 + \pi n), & x = \pi r, y = L_u \end{cases} \quad (4.8)$$

As can be seen in equation 4.8, L_d^{max} depends on the monitoring range L_u , radius of the pipe r , and order of helical path n as a multiplier of r . In practice, guided-waves are used for long ranges, up to hundreds of meters (Alleyne et al., 2001; Cawley et al., 2003; Davies et al., 2008). That is, practical monitoring ranges are significantly larger, by several orders of magnitude, than the radius of typical pipes in different applications ($L_u \gg r$). For example, the radius of the largest Schedule-40 pipe available is only about 0.3 m. If the monitoring range of this pipe is 100 m, L_u^2 is more than 110,000 times larger than r^2 . Therefore, terms including r^2 in the first three cases shown in equation 4.8 can be ignored. Note that the multiplier n needs to be very large (in this example, around 53) to compensate for such a huge difference between L_u^2 and r^2 . This number will be even larger for an smaller radius. I assume that, for damage detection purpose,

the arrivals from such large orders of helical paths can be ignored, since they will either be received after the sampling period or will be highly attenuated. After approximating the terms under square root in equation 4.8 with L_u , L_d^{max} depends on L_u and a product of radius r . Following the same logic, assuming that monitoring range is significantly larger than the pipe radius, L_d^{max} can be approximated with L_u , that is, $L_d^{max} \approx L_u$.

From the discussion above, it is concluded that $L_d^{min} \approx L_d^{max} \approx L_u$. In other words, for practical ranges of monitoring, regardless of the location of damage, the shortest damaged paths will be almost equal to the monitoring range.

The arrival times, however, depend on the velocity of the waves propagating to and from the damage. The wave velocity can be affected by the nature of the interference of the waves with damage (*e.g.*, mode conversion, velocity change, energy dissipation, *etc.*), which is mainly dictated by the type and geometry of the damage (Alleyne et al., 1998; Davies et al., 2008; Demma et al., 2004, 2003; Lowe et al., 1998a; Nurmalia et al., 2013). However, in this section, the type and geometry of the damage is considered to be constant and the only varying parameter between training and monitoring stages is the location of the damage. Therefore, the hypothesis to be tested in this section is as follows:

Hypothesis #4: For a particular type and geometry of damage, the difference between the location of the damage in training and monitoring stages does not significantly affect the detection performance of the SD method.

Detection Performance of Damage-sensitive Features

To test hypothesis #4 with high contingency, the *2-Loc1-big* (Figure 3.2c) and *2-Loc2-big* (Figure 3.2d) pipe layouts are used for training and testing, respectively. In the *2-Loc1-big* layout used for training, the shortest damaged path ($L_d^s + L_d^{h,0}$) is very close to L_u ($L_d = 1.07L_u$). However, in *2-Loc2-big* layout used for testing, the L_d/L_u ratio is unrealistically large ($L_d = 1.28 L_u$). This difference in the length of L_d will cause time-location of the scattered arrivals in training and test signals to be different, and hence, will adversely affect the detection performance of the trained coefficients. If hypothesis #4 is verified for this extreme case, the findings can be safely expanded to smaller differences in damage location, for damages of the same type and geometry.

A total of 30 training and 30 test datasets are created, with temperatures ranging from 25°C to 33°C. This translates to a total of 900 training/test combinations with different damage locations, and various ΔT s. Table 4.7 summarizes the average detection statistics for 1,800 scenarios, as a result of K-means clustering with 2-fold CV. As expected, the significant difference in the lengths of L_d s in the training and test pipes has slightly degraded the discriminatory power of damage-sensitive features compared to when only temperature (Table 4.1) or temperature and damage size (Table 4.4) vary.

Table 4.7: Average detection statistics for 900 training/test combinations trained and tested with different damage locations, and temperatures. Note: $25^\circ\text{C} \leq T_{tr} \text{ \& } T_{tst} \leq 33^\circ\text{C}$. Pipe layout *2-Loc1-big* is used for training, and layout *2-Loc2-big* is used for testing.

Average detection accuracy	Average FNR	Average FPR
91.0%	5.5%	11.0%

Online Damage Detection

Table 4.8 summarizes the performance of \hat{Y} s for online damage detection, using the coefficients trained by five different datasets, each for every temperature ranging from 25°C to 30°C. Each training dataset includes 200 intact observations and 200 observations from *2-Loc1-big* pipe layout (Figure 3.2c). Test dataset includes signals from *2-Loc2-big* pipe layout (Figure 3.2d), including 620 intact and 580 damaged observations. Despite the extreme scenario considered here regarding the difference in damage location ($L_d = 1.07L_u$ versus $L_d = 1.28L_u$), detection performance of the SD method remains satisfactory, which verifies hypothesis #4.

Table 4.8: Detection performance for online monitoring of pipes, using \hat{Y} s, when temperature, as well as the location of damage in training and test data is different. Note: *2-Loc1-big* pipe layout is used for training, and *2-Loc2-big* layout is used for testing. Observations are recorded in 1-minute intervals.

Training Temperature	Range of ΔT	Delay in Damage Detection (No. of observations)	Detection Accuracy
25°C	$0^\circ\text{C} \leq \Delta T \leq 7^\circ\text{C}$	11	99.0%
26°C	$-1^\circ\text{C} \leq \Delta T \leq 6^\circ\text{C}$	13	98.9%
28°C	$-3^\circ\text{C} \leq \Delta T \leq 4^\circ\text{C}$	15	98.7%
29°C	$-4^\circ\text{C} \leq \Delta T \leq 3^\circ\text{C}$	13	98.9%
30°C	$-5^\circ\text{C} \leq \Delta T \leq 2^\circ\text{C}$	25	97.9%
Average		15.4	98.7%

Evaluation of the Sparsity Assumption

Similar to previous experiments, the sparsity found to be weakly correlated with the location of damage in the training dataset. However, as can be inferred from Table 4.9, the average sparsity ratios found to be slightly lower for *2-Loc2-big* pipe layout, in which the damage is located

further from the transducers ($L_d = 1.28L_u$), as compared to the *2-Loc1-big* layout ($L_d = 1.07L_u$). All in all, in both cases, the solution of equation 4.2 remains sparse.

Table 4.9: Sparsity ratio (Sr, equation 4.4) statistics as a result of training SD algorithm with signals from a pipe with damage at two different locations, at different temperatures ($25^\circ\text{C} \leq T_{tr} \leq 33^\circ\text{C}$).

Pipe Layout	Average Sr.	Minimum Sr.	Maximum Sr.	Standard Deviation
<i>2-Loc1-big</i>	98.0%	88.0%	99.0%	3.0%
<i>2-Loc2-big</i>	82.0%	60.0%	99.0%	10.0%

4.4.4 Damage Size, Damage Location, and Temperature Variation

In this section, the discriminatory power of \hat{Y}_s is investigated when all three factors, namely temperature, damage size, and damage location, vary between training and test data. Table 4.10 summarizes the detection performance of the SD method for online detection implementation. Here, the coefficients are trained using data from *2-Loc2-big* pipe layout (Figure 3.2d, big damage and $L_d = 1.28L_u$) at temperatures ranging from 25°C to 30°C . Test data is randomly selected from a pool of observations recorded from *2-Loc1-sml* pipe layout (Figure 3.2c, small damage, and $L_d = 0.07L_u$), including 575 intact and 617 damaged observations. As reported in Table 4.10, presence of damage can be detected with high accuracies. However, it is notable that the change in the severity of damage could not be detected in the majority of the scenarios.

Table 4.10: Detection performance of \hat{Y} s for online monitoring of pipes, when temperature, damage size, and damage location in training and test data is different. Note: *2-Loc2-big* is used for training, and *2-Loc1-sml* is used for testing.

Training Temperature	Range of ΔT	Delay in Damage Detection (No. of observations)	Detection Accuracy
25°C	$0^{\circ}\text{C} \leq \Delta T \leq 7^{\circ}\text{C}$	19	98.4%
26°C	$-1^{\circ}\text{C} \leq \Delta T \leq 6^{\circ}\text{C}$	23	98.0%
28°C	$-3^{\circ}\text{C} \leq \Delta T \leq 4^{\circ}\text{C}$	52	95.6%
29°C	$-4^{\circ}\text{C} \leq \Delta T \leq 3^{\circ}\text{C}$	28	97.6%
31°C	$-6^{\circ}\text{C} \leq \Delta T \leq 1^{\circ}\text{C}$	19	98.4%
Average		28.2	97.6%

4.4.5 Physical Intuition about the SD Coefficients

Temperature Variation

High detection performance of the SD method for wide range of temperature variations implies that, for different temperatures, the time-location of the non-zero coefficients do not vary significantly. That is, for the temperature range that is considered in this study, and the pipe layout used in the experiments reported above, the time-locations of the sample points containing dominant damage information fall around the same region of the signals' time-trace. The change in the pipe layout, including location of transducers/damage and size of the damage, may affect the time-location of the coefficients, and therefore, affect the performance of the SD method. Effects of damage location and size of the coefficients will also be discussed. First, however, let us investigate the physical meaning of the extracted coefficients, for different temperatures.

Let x_i^j be the coefficient corresponding to the i th sample point in time ($i \in \{1, \dots, n\}$), when SD algorithm is trained with the j th dataset, at temperature T_j ($j \in \{1, \dots, N\}$). The average magnitude of the coefficients at each point of time i is obtained by equation 4.9:

$$\bar{x}_i = \frac{1}{N} \sum_{j=1}^N |x_i^j| : i \in \{1, \dots, n\} \wedge j \in \{1, \dots, N\} \quad (4.9)$$

Next, the \bar{x}_i is normalized by the standard deviation of $|x_i|$ values for all N training datasets (σ_i).

This results in standardized mean value of coefficients at point i :

$$\bar{x}_i' = \bar{x}_i / \sigma_i : i \in \{1, \dots, n\} \quad (4.10)$$

The SD algorithm (equation 4.2) is trained at temperatures ranging from 24°C to 32°C. For the experiments in this section, the *2-Loc2-big* layout (Figure 3.2d) is used. The transducers are located further from the two ends of the pipe segment (at one-third of the pipe length) in order to make the first arrival path significantly shorter than the end-reflection paths. In this layout, the shortest end-reflection path (A-D-A-B) is three times longer than the shortest undamaged path (A-B). This will help the first arrivals to be more distinguishable from end-reflections. In addition, the mass is located further from the transducers, *i.e.*, on the opposite side of the pipe, so that the first arrivals from the undamaged and damaged paths can be better distinguished. Considering the dispersion curve of this pipe layout, the first arrival from damaged path (A-C-B) should be received at about 50 μ s after the last arrival from undamaged path (A-B).

Figure 4.6 depicts \bar{x}'_i values (equation 4.10) for a total of $N=30$ training datasets ($24^\circ\text{C} \leq T_{tr} \leq 32^\circ\text{C}$), from *2-Loc2-big* pipe layout. PCDisp (Seco and Jiménez, 2012), an open source software for modeling guided-waves in cylindrical media, is used in this thesis to calculate the first arrival times shown in Figure 4.6.

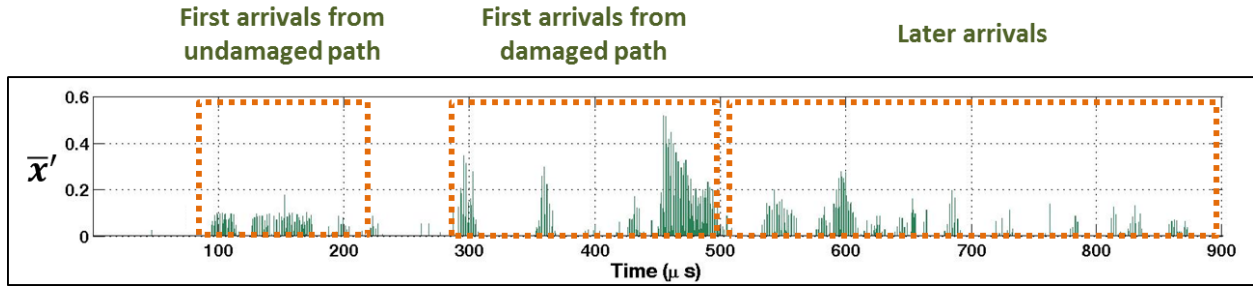


Figure 4.6: Mean of the coefficients at all sample points in time, normalized with the standard deviation of the magnitudes (equation 4.10), for $24^\circ\text{C} \leq T_{tr} \leq 32^\circ\text{C}$. Note: First arrival times are calculated for *2-Loc2-big* pipe using PCDisp (Seco and Jiménez, 2012).

Expectedly, the first arrivals from undamaged path are associated with smaller magnitudes of coefficients and/or with larger variance (smaller \bar{x}'_i 's). It can also be seen that the first arrivals from the shortest damaged path are associated with the largest coefficients, and/or smallest variance, for all ranges of training temperatures. These findings are in agreement with the reasoning behind development of the SD method. That is, the arrivals that have illuminated the damaged section of the pipe may include more significant damage information. Interestingly, several sample points consisting of later arrivals are also associated with large coefficients, indicating their importance in defining the state of the pipe. Denoising methods that are based on removing multi-path reflections should be applied with caution since, as can be seen in Figure 4.6, multi-path reflections and later arrivals can contain significant damage information.

Damage Size and Temperature Variation

In this section, the SD coefficients trained with different sizes of damage are compared. Similar to previous experiments, in order to better distinguish between the first and later arrivals, *2-Loc2-sml* and *2-Loc2-big* pipe layouts are used (Figure 3.2d). The SD algorithm is trained with signals captured from the pipe exposed to two different sizes masses, at temperatures ranging from 24°C to 33°C. Using equation 4.9, average coefficients are obtained for both cases, namely \bar{x}_{sml} and \bar{x}_{big} . Cross-correlation between \bar{x}_{sml} and \bar{x}_{big} can be used to compare the time-location of the coefficients of the two cases. Figure 4.7 shows that the lag corresponding to the maximum correlation between \bar{x}_{sml} and \bar{x}_{big} is zero. In other words, the peak values of the coefficients corresponding to both sizes of mass happen at, roughly, the same points in time.

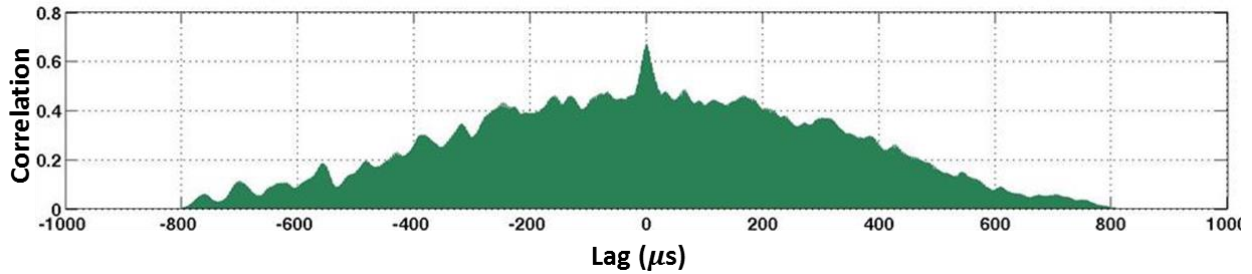


Figure 4.7: Cross-correlation between the mean of the coefficients trained with small damage (2-Loc2-sml) and big damage (2-Loc2-big), at $24^{\circ}\text{C} \leq T_{tr} \leq 33^{\circ}\text{C}$.

Damage Location and Temperature Variation

In this section, the SD coefficients trained with different locations of damage are compared. The *2-Loc1-big* and *2-Loc2-big* pipe layouts are used (Figure 3.2c and 3.2d). The SD algorithm is trained for both locations, with temperatures varying between 24°C and 33°C. Using equation 4.9, average coefficients are obtained for both damage locations, namely \bar{x}_{loc1} and \bar{x}_{loc2} . Figure 4.8

shows the cross-correlation between \bar{x}_{loc1} and \bar{x}_{loc2} . It can be seen that the lag corresponding to the maximum correlation among \bar{x}_{loc1} and \bar{x}_{loc2} is zero. That is, the peak values of the coefficients, trained with either of the damage locations, are generally located at the same points in time, even for the extreme location difference that is examined here.

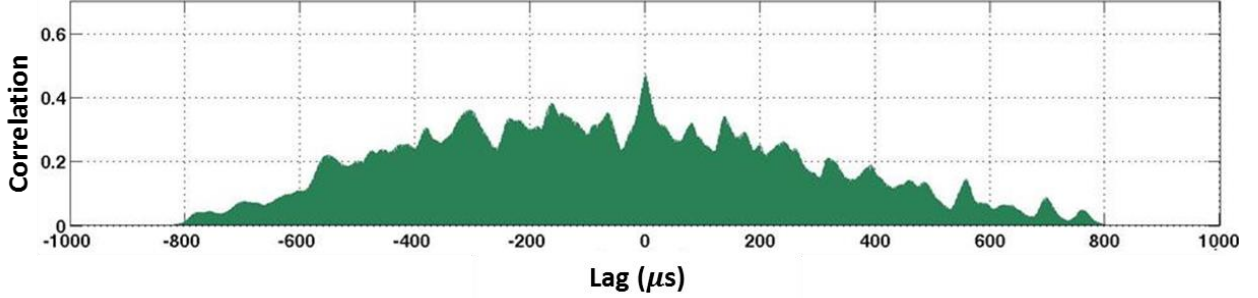


Figure 4.8: Cross-correlation between the mean of the coefficients trained for two different locations of damage (2-Loc1-big and 2-Loc2-big), at $24^{\circ}\text{C} \leq T_{tr} \leq 33^{\circ}\text{C}$.

4.5 Validation of the SD Method

4.5.1 Validate the Sparsity Assumption with Field Data

In this section, I verify the sparsity assumption for addressing the curse of dimensionality brought about in equation 4.2. That is, I assume that a sparse subset of a guided-wave signal contains enough damage information for optimal damage detection.

The objective is to verify whether damage detection performance increases as the signals become sparser. For this purpose, the feature extraction method proposed by Liu et al (2013) is used,

because it has shown high potential for damage detection of pipes under varying EOCs. In their proposed approach, Singular Value Decomposition (SVD) is applied on guided-waves, and the projections of the signals on one of the left singular vectors is used as the damage-sensitive feature (for more details please refer to Liu et al. (2013)).

A subset of the field data described in Section 3.3 is used here. Theoretically, increasing the value of the regularization parameter (ξ) in equation 4.2 will increase the sparsity of the solution. The SD algorithm is trained for different values of the regularization parameter ξ , with 0.01 increments. Since the results do not vary significantly for very large values of ξ , Figure 4.9 only reports the findings for $0 \leq \xi \leq 1$. The solid red line in Figure 4.9 shows the sparsity ratios (equation 4.4) at different values of ξ .

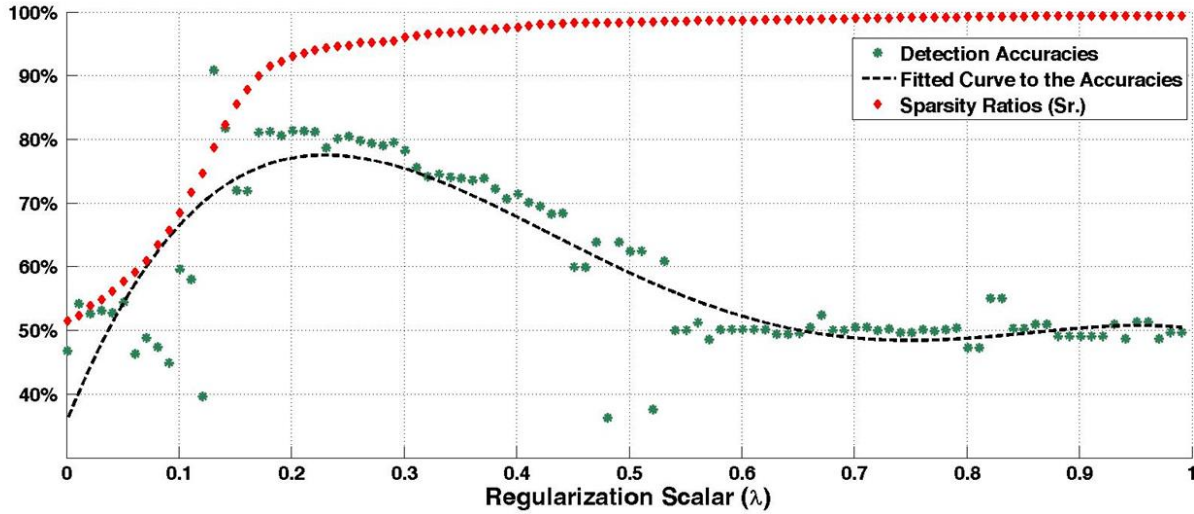


Figure 4.9: Variation of sparsity ratio (solid red line) and detection accuracy of the SVD method (Liu et al., 2013) (dashed black line) for different values of regularization scalar ξ in equation 4.2.

It can be seen that for ξ values of up to around 0.2, a sharp increase in the sparsity is observed. After this point, the change in the sparsity becomes more gradual. This behavior implies that, after a certain point, increasing the sparsity while retaining enough damage information for prediction of class labels becomes more difficult.

Sparse test signals are obtained using the coefficients trained with different ξ s. At each ξ , sparse test signals are the subset of the signals that correspond to non-zero coefficients at that ξ . The dashed black line in Figure 4.9 shows the detection accuracy of the SVD method when applied to signals with different sparsities. As can be seen in this figure, applying the SVD method to the complete time-trace of signals (*i.e.*, zero sparsity, $\xi = 0$), leads to a detection accuracy of around 47%. As we increase the sparsity of the test signals, detection accuracies improve, and reach their maximum when the sparsity ratios are around 90% to 97%. This trend verifies the sparsity assumption of this study. That is, there is a sparse subset of the signals that can lead to optimal damage detection under varying EOCs. It is also worth noting that further increasing the sparsity leads to a decrease in detection accuracy until it again hits the accuracy range of zero-sparsity, or the case that classification is purely random. This pattern indicates that, similar to noisy non-sparse signals, too sparse solutions can degrade the performance, because of losing damage information.

4.5.2 Online Damage Detection under Multiple EOCs

In the laboratory experiments reported throughout this chapter, temperature is the only EOC that is significantly varying during the data collection process. In this section, high performance of

the SD method for online damage detection is validated using the signals recorded under a wide range of varying EOCs. Figure 4.10 illustrates an example of online monitoring implementation for a subset of field data explained in Section 3.3. Figure 4.10a and 4.10b show the abrupt variation of temperature and fluid flow rate, respectively, among other EOCs discussed in Section 3.3. The sharp drop in the predicted labels of the test observations (Figure 4.10c) indicates the time when damage (mass scatterer) is introduced.

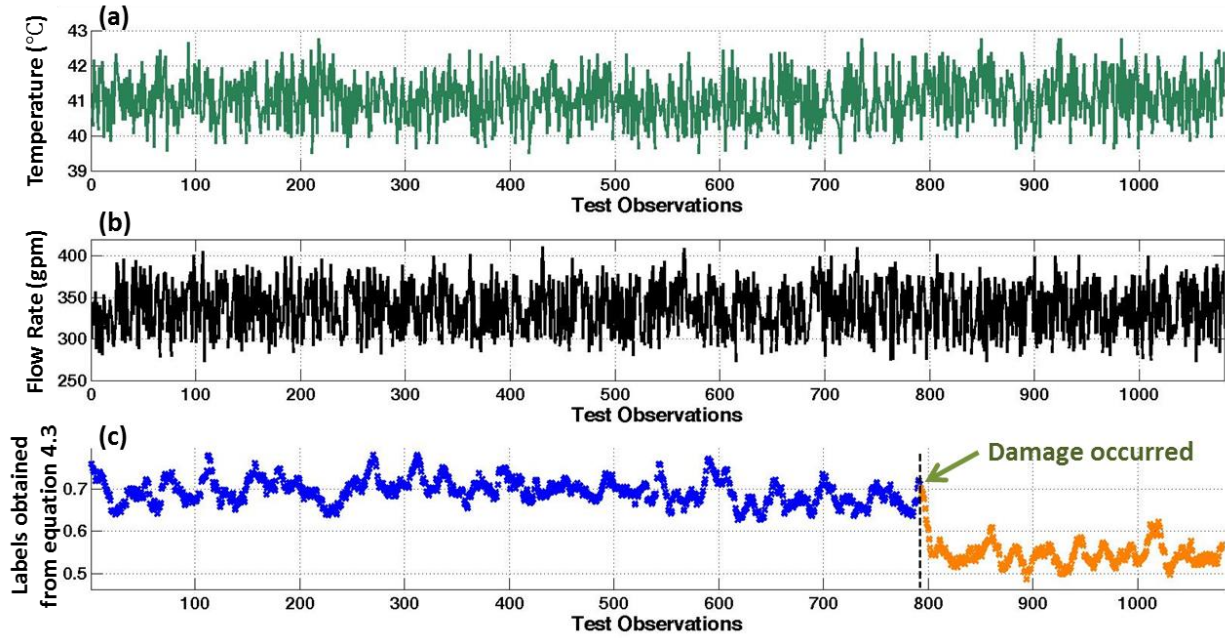


Figure 4.10: Online damage detection using \hat{Y}_j s for filed data explained in Section 3.3: (a) Temperature variation of the test observations, (b) Fluid flow rate variation of the test observations, (c) Predicted class labels (\hat{Y}_j s) for test observations.

Table 4.11 summarizes online detection performance of the SD method, as the average of 10-fold CV, for data recorded from four different summer days and three winter days. The simple detection algorithm explained in Section 4.4.1 is used here. That is, when the distance between the label of the j th observation (\hat{Y}_j) and the average of the labels in the window before the j th

observation (\hat{Y}_{j-20} to \hat{Y}_{j-1}) is larger than ten times the standard deviation of the labels in the window, occurrence of damage is detected ($\Delta\hat{Y} > 10\sigma$). This table also summarizes the average sparsity ratios for all the iterations of 10-fold CV. The SD method improves the performance of the singular value decomposition (SVD) method reported by Liu et al. (2013) which is applied on part of the data reported in this section.

Table 4.11: Average online detection performance and sparsity ratios, for 10-fold CV, using field data explained in Section 3.3, for four summer days and three winter days. Observations are recorded in 1-minute intervals.

Data	Average Delay in Detection (No. of observations)	Average Detection Accuracy	Average FNR	Average Sparsity Ratio (Sr.)
Aug 9 th	3.8	99.5%	0.5%	98.0%
Aug 10 th	2.8	99.8%	0.3%	97.0%
Aug 24 th	5.2	96.2%	7.0%	99.0%
Aug 31 st	4.0	99.6%	1.3%	98.7%
Jan 18 th	2.2	99.8%	1.5%	98.6%
Jan 19 th	2.6	99.7%	2.0%	99.6%
Jan 23 rd	1.8	99.8%	0.2%	99.0%
Average	3.2	99.2%	1.8%	98.6%

4.5.3 Detection of a Small Crack and Mass Loss

It is important to evaluate the performance of the SD method for detection of damages of different type from training stage. As discussed in Chapter 1, arrival times can be affected by the nature of the interference of the waves with damage (*e.g.*, mode conversion, velocity change, energy dissipation, *etc.*), which is mainly dictated by the type and geometry of the damage

(Alleyne et al., 1998; Davies et al., 2008; Demma et al., 2004, 2003; Lowe et al., 1998a; Nurmalia et al., 2013). In this section, the proposed SD method is evaluated for detection of actual damages of different type from the mass scatterer that is used in training.

A small crack is simulated by a saw-cut on the surface of the steel pipe explained in Section 3.2. Later, a small mass loss is introduced to the pipe, in a different location than the crack. For online monitoring applications, evaluating the SD method for detection of consecutive damages is important. As discussed in Section 4.4.1, in this work, I suggest a non-adaptive approach in which the training dataset is not updated as new signals are obtained during the monitoring stage. Therefore, failing in detecting the first damage does not affect detection of the second damage by degrading the accuracy of the trained discriminant vector. However, the question is whether the method can still distinguish between damaged and undamaged pipe when multiple damages occur at the same time, or ideally, can detect the change in the state of the pipe when the second damage occurs.

For these experiments, a Schedule-40 steel pipe (Figure 3.3) with 33.5 *mm* outer diameter and 3.3 *mm* thickness is used. Before introducing the crack, ultrasonic pitch-catch records are obtained from the intact pipe and from the pipe with a small mass scatterer placed in a non-symmetric location shown in Figure 3.3, at different temperatures ranging from 25°C to 27°C. These signals are used for training. Later, an oblique cut of 2.5 *cm* long, extended in both circumferential and longitudinal directions of the pipe (see Figure 3.3), is imposed at the middle of the length of the pipe, using a jewelers saw. The maximum thickness along the length of the crack is approximately 0.34 *mm* (10% of the pipe thickness), and the maximum depth is

approximately 0.7 mm (20% of the pipe thickness). The second damage is a small mass loss (Figure 3.3) with 5 mm diameter and maximum depth of 0.25 mm ($\sim 8\%$ of the pipe thickness). The location of the second damage differs from both the crack, and training mass scatterer location. For testing, signals from the intact pipe and the pipe with crack, and with both crack and mass loss are used at $24^\circ\text{C} \leq T_{tst} \leq 32^\circ\text{C}$.

In this set of experiments, in addition to the range of temperature variation, the location, size, number and the type of the structural abnormality in the training and test data are different, which strengthens the generality of the results. Figure 4.11 shows that not only can the proposed SD method can distinguish damaged and undamaged pipes when more than one damage present, but it may also detect the change in the severity of the damage and/or occurrence of subsequent damages.

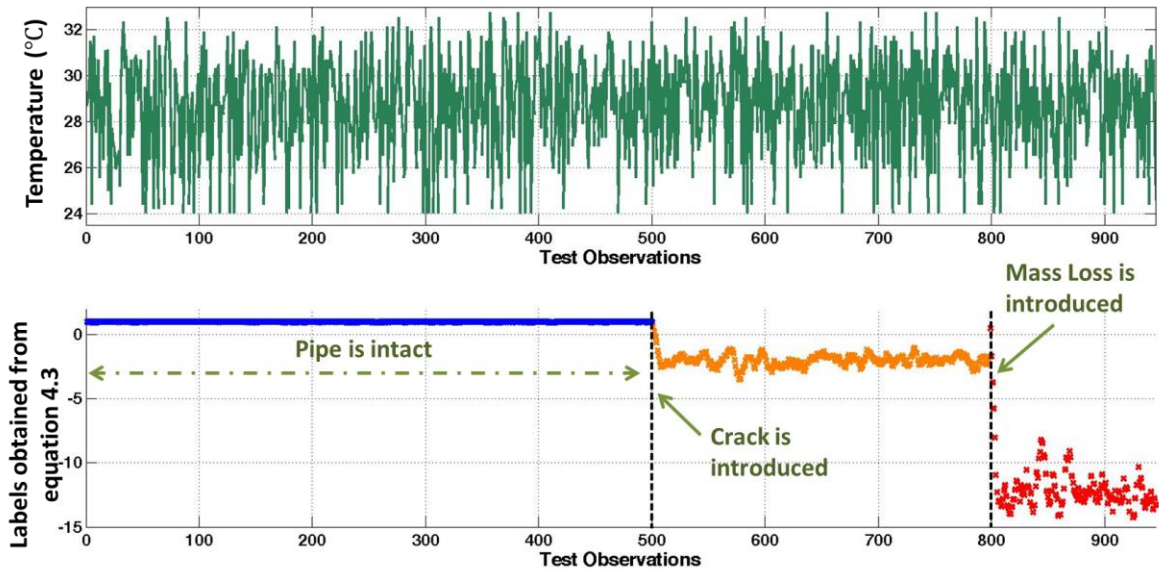


Figure 4.11: Online detection of a 2.5 cm long oblique crack (with maximum thickness of 0.34 mm and maximum depth of 0.7 mm), and a subsequent small mass loss (with maximum depth of 0.25 mm), in a Schedule-40 steel pipe.

4.5.4 Non-Gaussianity of Guided-waves

Linear discriminant analysis (LDA) is a common method in pattern recognition for extracting a subspace of the variables that discriminates between the observations of different classes (*i.e.*, discriminant vector). However, LDA assumes that the variables are normally distributed. Therefore, in order to apply LDA in the time-domain of guided-waves, n sample points in the signals have to follow a Gaussian distribution. To evaluate Gaussianity, two common statistical tests, namely *Chi-square goodness-of-fit* (uses the variables' mean and variance) and *Jarque-Bera* test (without the knowledge of mean and variance) are used. These two metrics evaluate the null hypothesis that the variables (sample points in time) come from a normal distribution. The dataset includes 5,500 signals from the intact pipe, and 5,500 signals from the damaged pipe, at temperatures ranging from 24°C to 38°C. The pipe layout *1-loc1-sml*, shown in Figure 3.2a, is used for these experiments. Table 4.12 summarizes the percentage of the sample points in time, after the first arrivals, for which the null hypothesis is rejected at 5% significance level. As reported in this table, almost none of the sample points in these signals follow a normal distribution. This means that LDA cannot be applied to guided-waves in time-domain.

Table 4.12: Percentage of the sample points, after the first arrivals, for which the null hypothesis of Gaussianity is rejected at 5% significance level.

Gaussianity Test	Signals from Intact Pipe	Signals from Damaged Pipe
<i>Chi-square goodness-of-fit</i>	100.0%	100.0%
<i>Jarque-Bera</i>	99.8%	99.7%

Chapter 5

An Unsupervised Approach for Damage Detection of Pipelines

5.1 Introduction

The SD method discussed in Chapter 4 addresses the challenges derived from complex nature of guided-waves and their sensitivity to EOC variations. Being a supervised approach, its performance may be affected by the differences between training and test signals. EOCs, damage characteristics and damage location are some of the parameters that may affect the detection performance of the SD method. The robustness of the method to these parameters is validated in Chapter 4 with extensive set of laboratory and field experiments. These findings suggest that a simple binary labeled training data (*i.e.*, damaged versus undamaged), obtained under a limited range of EOCs, is enough for optimal performance of the SD method.

However, the application of the SD method is limited to the cases where obtaining the training data from a pipe with abnormality is possible and practical. Moreover, it is useful to reduce the number of training parameters that may affect the performance of a detection method. Therefore, to address such issues, the objective of this chapter is to develop an *unsupervised* method for damage detection in pipes; specifically, one that can also deal with the complexity of guided-waves and their sensitivity to EOC variations. In the following sections of this chapter, research motivations for the proposed unsupervised method are discussed. Next, research questions to be addressed are listed. An unsupervised method and the corresponding framework for continuous damage detection are proposed. The answer to the research questions are experimentally investigated. Finally, the performance of the proposed method is validated using both laboratory and field data.

5.2 Motivation

The unsupervised approach that will be proposed in this chapter is motivated by the nature of the damage effects on the signals, and thus on their energy content. Damage, either in the form of variation in wall thickness or change in surface roughness, can alter propagation of guided-waves in a pipe. These changes can be in the form of mode conversion, energy dissipation, change in phase and/or group velocities, modal attenuation, energy exchanges, *etc.* (Alleyne and Cawley, 1992; Cawley et al., 2002; Davies et al., 2008; Long et al., 2003b; Løvstad and Cawley, 2012; Lowe et al., 1998b; Nurmalia et al., 2013). As discussed in Chapter 1, an intuitive method to detect these changes is to simply subtract the signal from a baseline signal captured from the undamaged pipe. Ideally, the changes caused by damage will result in an increase in the

amplitude of the residual. However, as discussed in Section 1.3.3, similar to damage, EOCs alter guided-wave signal, and therefore, may mask and/or appear as the effects of damage, leading to type I and type II errors (Croxford et al., 2010; Degtyar et al., 1996; Long et al., 2003a; Raghavan and Cesnik, 2007, 2007; Scalea and Salamone, 2008). Studies such as Clarke et al. (2009), and Konstantinidis et al. (2007) have shown that after compensating for the effects of temperature, the amplitude difference between the two signals is a good indicator of damage existence. However, these compensation methods may only be effective for small variations of temperature (± 0.5 to 1.0 °C), depending on the complexity of the structure and the number of propagating modes (Eybpoosh et al., 2015; Clarke et al., 2009; Lu and Michaels, 2005). In other words, although damage manifests itself as changes in the signals, detecting such changes in complex signals captured under realistic conditions is not trivial in practice.

Let us refer again to a schematic depiction of wave propagation in an unrolled view of a pipe shown in Figure 5.1. Let point A be the transducer and point B be the receiver in a pitch-catch setup. The excited guided-waves travel in multiple paths, and depending on the modal properties, physical/geometric features of the pipe, and EOCs, different portions of these waves will arrive to the sensor B at different points in time during the sampling period. Among the recorded samples, the effects of EOCs will be more significant on low-energy arrivals with small signal-to-noise ratios (SNR) compared to high-energy arrivals. For example, from the experiments conducted by Raghavan and Cesnik (2007) and Yeum et al. (2014), it can be inferred that the amplitude and time location of the arrivals from S0 mode with larger energy content are less affected by the temperature variation as opposed to latter arrivals of A0 mode with smaller energy. In terms of complex diffuse-field guided-waves, in which different wave modes are not

easily distinguishable from each other, studies such as Lu and Michaels (2005) and Eybpoosh et al. (2015) have shown that the effects of temperature on later arrivals, potentially with lower energy content, are more significant than initial arrivals.

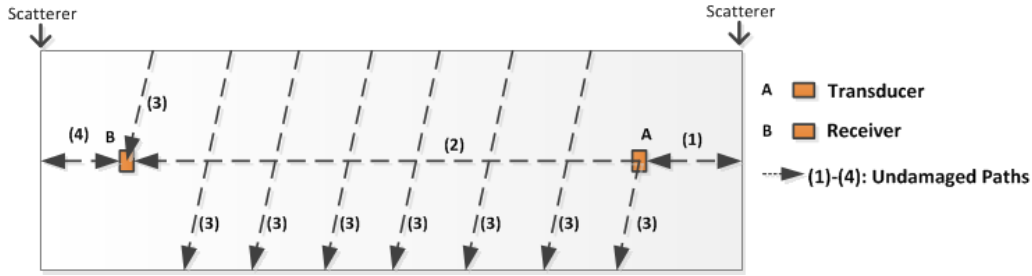


Figure 5.1: A schematic unrolled view of a pipe, illustrating example wave propagation paths in an undamaged pipe.

Therefore, a hypothesis based on these observations is that, the effects of EOCs will be less dominant in a sparse subset of arrivals with higher energy content, as opposed to noisy complex diffuse-field signals. If this hypothesis is true, such a subset may be used for detecting the effects of damage. To be an unsupervised method, however, extracting such a subset should not require information about the state of the pipe.

5.3 Method

The goal is to develop an unsupervised method to extract a subset of the guided-wave signals that contains the majority of the signal's energy. The optimal solution will be the “simplest” subset with minimum sensitivity to EOC variations. The main hypotheses of the proposed method are as follow:

Hypothesis #1: A sparse subset of the arrivals in a diffuse-field signal represents the majority of the signal's energy.

Hypothesis #2: The EOC effects are less dominant in a sparse subset of the arrivals that contains the majority of a diffuse-field signal's energy, compared to the complete signal.

These hypotheses will be addressed through research questions #1 and #3 presented in Section 5.3.1, and will be verified through a variety of experiments reported later in this chapter. Research question #2 will evaluate whether the aforementioned sparse subset in Hypothesis #2 can be used for damage detection or not. An unsupervised method is proposed to obtain a sparse representation of signals so that the signal energy can be optimally reconstructed. It is expected that such a sparse representation be mainly composed of high-energy arrivals. For this sparse energy method (which will be referred to as the SE method), the training dataset is not labeled, but assumed to be mostly composed of signals captured from an intact pipe. Section 5.5.4 evaluates the performance of the proposed method as the number of damaged signals in the training dataset increases.

5.3.1 Research Questions

Based on the discussions in the motivation section, and the hypotheses given above, following is the list of research questions that need to be answered:

Question #1: Does a sparse subset of the arrivals in a diffuse-field guided-wave signal from a pipe (with the transducer setup and specifications defined in Chapter 2 of this thesis) contain the majority of the signal's energy?

Question #1.1: How does the sparsity of this subset relate to the amount of energy it represents?

Question #1.2: How can a sufficient level of sparsity be selected for optimal damage detection?

Question #2: How well, as measured by detection accuracy, false positive rate (FPR) and false negative rate (FNR), can the effects of damage be detected by using the extracted subset?

Question #3: How sensitive are the extracted subset and any damage-sensitive feature used in Question #2 to the variations of EOCs specified in Section 2.2 of this manuscript?

5.3.2 Overview of the Sparse Energy (SE) Method

Proposed SHM Framework based on the SE Method

Figure 5.2 shows the proposed SHM framework for continuous damage detection of pipelines using the method explained in this chapter. Different components of the two training and monitoring stages will be explained in the next section, and will be evaluated experimentally throughout this chapter.

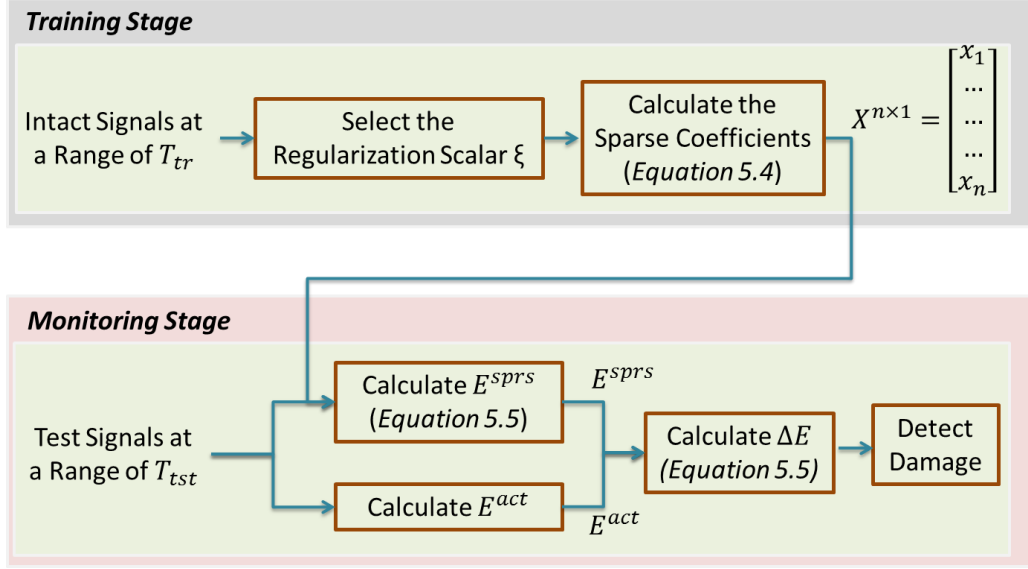


Figure 5.2: Application of the SE method for continuous monitoring of pipelines.

Training Stage

Let m_{tr} be the total number of training signals, and n be the length of the signal in time (*i.e.*, the number of sample points in the time-domain). Let $A^{m_{tr} \times n}$ (equation 5.1) be the matrix of element-wise squared values of training signals (assumed to be mostly from an intact pipe). In other words, the i th element in the j th row of matrix $A^{m_{tr} \times n}$ ($d_{i,j}^2, i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m_{tr}\}$) is the energy of the i th arrival of the j th signal.

$$A^{m_{tr} \times n} = \begin{bmatrix} d_{1,1}^2 & d_{1,2}^2 & \dots & d_{1,n}^2 \\ \dots & \dots & \dots & \dots \\ d_{m_{tr},1}^2 & d_{m_{tr},2}^2 & \dots & d_{m_{tr},n}^2 \end{bmatrix} \quad (5.1)$$

Let $\vec{E}^{m_{tr} \times 1}$ be the vector of training signals' energy (equation 5.2). That is, the j th element in $\vec{E}^{m_{tr} \times 1}$ vector ($E_j, \forall j \in \{1, \dots, m_{tr}\}$) is the energy of the j th signal, or projection of $A^{m_{tr} \times n}$ into the vector of ones ($\vec{1}$).

$$\vec{E}^{m_{tr} \times 1} = \begin{bmatrix} E_1 = \sum_{i=1}^{i=n} d_{1,i}^2 \\ \dots \\ E_{m_{tr}} = \sum_{i=1}^{i=n} d_{m_{tr},i}^2 \end{bmatrix} = A^{m_{tr} \times n} \vec{\mathbf{1}} \quad (5.2)$$

The goal is to obtain a sparse vector of coefficients $\vec{X}^{n \times 1}$ so that the projection of $A^{m_{tr} \times n}$ onto this vector results in good reconstruction of signal energies in $\vec{E}^{m_{tr} \times 1}$. That is, the sum of the energy residuals (*i.e.*, the difference between the actual energy and the energy obtained through projection) for all training signals ($\vec{d}_j^{1 \times n} \forall j \in \{1, \dots, m_{tr}\}$) approaches zero (equation 5.3). Here, it is worth reemphasizing that the training signals are not labeled (making it an unsupervised approach), but are assumed to mostly include intact signals (see Section 5.5.5).

$$\sum_{j=1}^{j=m_{tr}} |\vec{d}_j^2 \vec{X} - \vec{d}_j^2 \vec{\mathbf{1}}| \approx \mathbf{0} \quad (5.3)$$

In order to impose sparsity, similar to the method discussed in Chapter 4, a regularization scalar ξ is introduced to penalize the arrivals that have less contribution in estimating the signal energy. Equation 5.4 summarizes the proposed algorithm for extracting the vector of coefficients $\vec{X}^{n \times 1}$. The ℓ_1 norm in the first part of equation 5.4 minimizes the sum of the residuals shown in equation 5.3, and the ℓ_1 norm of the $\vec{X}^{n \times 1}$ vector enforces the sparsity by assigning close-to-zero coefficients to the arrivals with smaller contribution in reconstructing the energy.

$$\underset{x}{\operatorname{argmin}} \left\{ \|A^{m_{tr} \times n}(\vec{X}^{n \times 1} - \vec{\mathbf{1}})\|_1 + \xi \|\vec{X}^{n \times 1}\|_1 \right\} \quad (5.4)$$

Monitoring Stage

Damage interference can alter guided-wave signals. These effects, either in the form of the change in wave velocity, or energy dissipation and/or mode conversion, may change the location, in time, where the high-energy components appear in the signal. In that case, the coefficients obtained from the intact signals in equation 5.4, will not be a good representative of high-energy arrivals in a damaged signal.

Equation 5.5 calculates the energy residual of the j th signal of the monitoring dataset ($\Delta E_j, \forall j \in \{1, \dots, m_{tst}\}$). In other words, ΔE_j is the difference between the actual energy of the signal (E_j) and the energy of the sparse signal (E_j^{sprs}), estimated by projecting the j th squared signal (\vec{d}_j^2) onto the \vec{X} vector.

$$\Delta E_j = \frac{|\vec{d}_j^{2T} \vec{X} - \vec{d}_j^2 \vec{1}|}{\vec{d}_j^2 \vec{1}} = \frac{|E_j^{sprs} - E_j|}{E_j} \quad (5.5)$$

The coefficients in the \vec{X} vector are obtained using mostly the *intact* signals. Therefore, the coefficients with larger magnitudes correspond to the arrivals, in an *intact* signal, with larger contribution in reconstructing the signal energy. Therefore, if the monitored j th signal is from an intact pipe, the residual ΔE_j in equation 5.5 is expected to be close to zero. However, if the j th signal is from a damaged pipe, the value of the residual ΔE_j is expected to increase. If so, energy residuals (ΔE) can be used as *damage-sensitive feature* for detection purpose. Therefore, the third hypothesis to be evaluated in this chapter is as follows:

Hypothesis #3: Due to the effects of damage, the extracted sparse vector of coefficients from training intact signals will not be a good representative of the energy of damaged signals, which will lead to an increase in the magnitude of the energy residuals (ΔE in equation 5.5).

5.4 Experimental Investigation of the SE Method

This section provides a general proof of the concept for the proposed SE method. The basic assumptions of the SE method, research questions, and different aspects of the framework proposed in Figure 5.2 are studied using the experimental data captured from the aluminum pipe discussed in Section 3.1.

5.4.1 Sparse Representation of Signals for Energy Estimation (Question #1)

The objective of this section is (1) to investigate whether a sparse representation of diffuse-field guided-wave is sufficient to represent the majority of the total energy of the signal (*i.e.*, hypothesis #1 and research question #1), and (2) to evaluate the relationship between the level of sparsity and the error in estimating the signals' energy (*i.e.*, research question #1.1).

Pitch-catch records from the intact state of the aluminum pipe *1-loc1-sml* (Figure 3.2a) are used for training. The temperature of the training signals varies between 24°C and 38°C. The SE algorithm given in equation 5.4 is trained with different values of regularization scalars (ξ). For simplicity, the ξ increments increase as the values of ξ get larger. These increments are listed in Figure 5.3.

Based on the SE algorithm given in equation 5.4, increasing the regularization scalar ξ will enforce sparser solutions. At any given ξ , that is, at any level of sparsity, the average training error (ΔE) is used as a metric to evaluate the accuracy of energy reconstruction. In equation 5.6, $\overline{\Delta E}_\xi$ is the average training error corresponding to regularization scalar ξ , m_{tr} is the total number of training signals, E_j^{act} is the energy of the j th signal in the training dataset, and $E_j^{sprs(\xi)}$ is the energy of the sparse j th signal corresponding to regularization scalar ξ .

$$\overline{\Delta E}_\xi = \frac{1}{m_{tr}} \sum_{j=1}^{j=m_{tr}} \frac{|E_j^{act} - E_j^{sprs(\xi)}|}{E_j^{act}} \quad (5.6)$$

To evaluate the sparsity of the solution in each round, two metrics are considered: (1) the ℓ_1 norm of the \vec{X} vector, and (2) a sparsity ratio as is defined in equation 5.7. The ℓ_1 norm is the sum of the absolute values of the coefficients in the \vec{X} vector. The sparser the \vec{X} vector, the smaller the ℓ_1 norm. Sparsity ratio is a metric to measure the ratio of zero coefficients to the total length of the signal. It is notable that by using the ℓ_1 norm of the \vec{X} vector as a regularizer in the SE algorithm (equation 5.4), the value of the coefficients may never be absolute zero, but rather be very small. To handle this, any coefficient whose magnitude is smaller than the largest coefficient assigned to the initial part of the signal (*i.e.*, before the first arrivals), when $\xi = 0$, is considered to be zero. This is because the energy content of this part of the signal is expected to have minimal contribution to the total energy of the signal. In equation 5.7, n is the number of sample points in time trace of the signal, x_i is the coefficient assigned to the i th sample point in the signal, and $x_{noise}^{max}(\xi)$ is the largest coefficient of the initial part of the signal corresponding to

regularization scalar ξ . $Sr(\xi)$ in equation 5.7 is the sparsity ratio corresponding to the regularization scalar ξ .

$$Sr(\xi) = \frac{|\{i \in \{1, 2, \dots, n\} : |x_i| \leq |x_{noise}^{max}(\xi)|\}|}{n} \times 100 \quad (5.7)$$

Figure 5.3a shows the variation of training error $\overline{\Delta E}_\xi$ with respect to regularization scalar ξ . As can be seen in this figure, $\overline{\Delta E}_\xi$ remains close to zero for ξ up to around 1.0. This range corresponds to the sparsity ratios as large as 97-99% (Figure 5.3b). It is important to note that in these experiments, 300 training signals ($m_{tr} = 300$) with length of 9000 sample points ($n = 9000$) are used. For very small values of ξ , the solutions of the algorithm will assign close-to-zero coefficients to at least 8,700 ($n - m_{tr}$) of the sample points, and non-zero coefficients to the rest. This trend will go on until the ξ values for which such solutions do not satisfy the objective. The sparsity ratio corresponding to 300 non-zero coefficients is about 97%. As can be seen in Figure 5.3b, this corresponds to $0.001 \leq \xi \leq \sim 0.01$. Therefore, true sparse solutions correspond to the ξ values greater than 0.01, and of course smaller than 1.0, considering the training errors. Figure 5.3a shows the zoomed depiction of the training errors in this range. It can be seen that for sparsity ratios as large as 98%-99%, training errors as small as $0.5 \times 10^{-3} - 4 \times 10^{-3}$ can be obtained. In other words, with only 1%-2% of the arrivals in the time trace of the signal, the energy of the signal can be estimated with above 99% accuracies. This suggests that a sparse subset of the arrivals in a diffuse-field guided-wave signal is sufficient to estimate its total energy, and hence verifies hypothesis #1 and research question #1 discussed in Section 5.3.

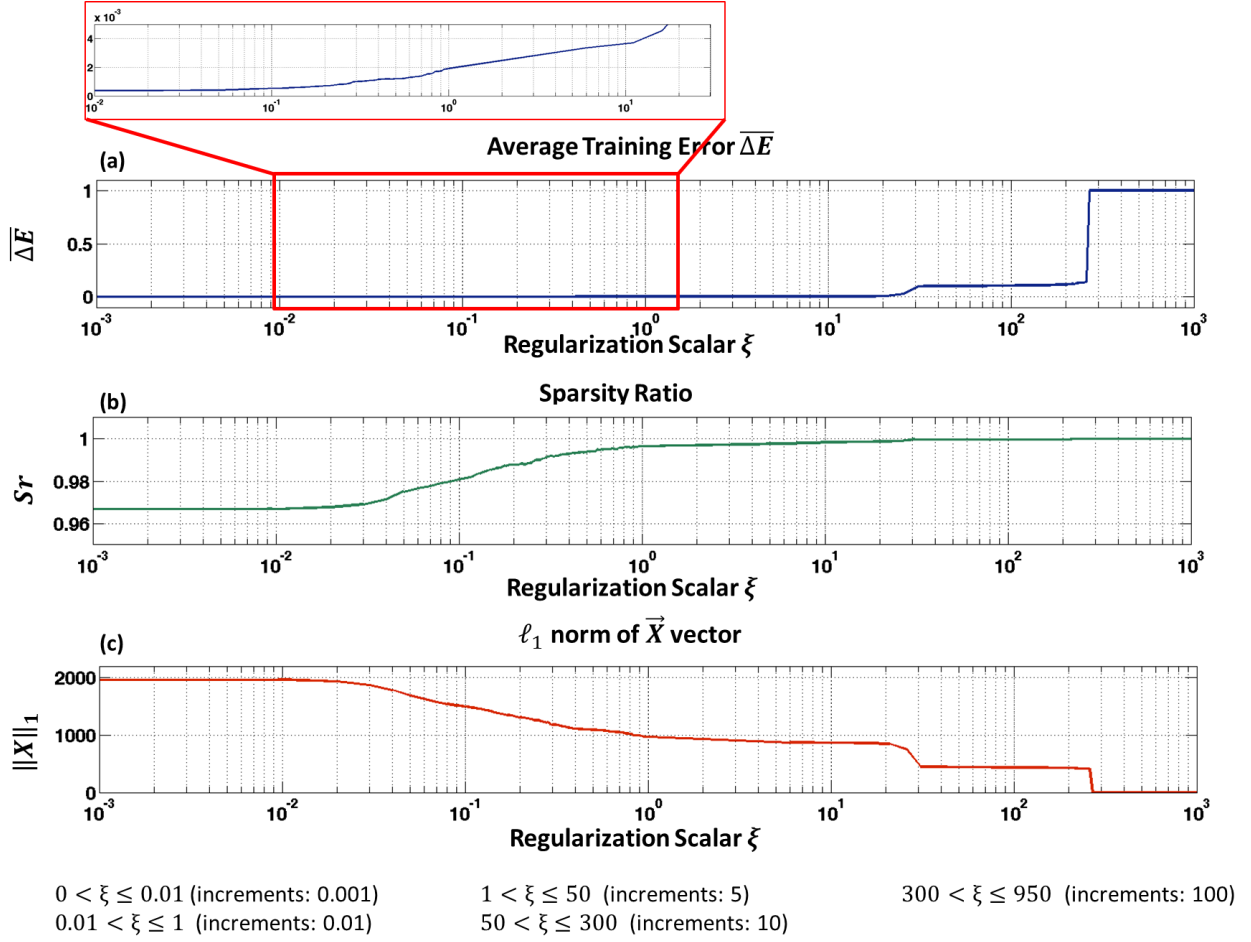


Figure 5.3: Variation of (a) average training error (equation 5.6), (b) sparsity ratio (equation 5.7), and (c) ℓ_1 norm of \vec{X} vector, with respect to regularization scalar ξ . Note: Regularization scalars are presented in logarithmic scale.

Figure 5.3, however, shows that increasing the sparsity more than a certain limit increases the training error significantly, reaching close to 100% for large values of ξ . This suggests that enforcing too sparse solutions will result in eliminating the arrivals that contain significant portion of the signal's energy. Therefore, in order to select a suitable ξ , a tradeoff between the sparsity and estimation accuracy is needed. Section 5.4.7 summarizes an approach for selecting the optimal regularization scalar ξ by trading off between simplicity, energy estimation accuracy, and robustness to EOCs. These discussions will address the second half of question #1.

5.4.2 Stability of the SE Method

Before moving forward, in order to ensure the repeatability of the findings, it is important to determine whether the SE algorithm produces unique solutions. For the experiments reported in this chapter, the Matlab convex optimization package (*i.e.*, *cvx* by Grant and Boyd (2008)) is used to solve the Lasso optimization problem formulated in equation 5.4. This algorithm chooses random initial values at each run, therefore, it is important to ensure that the solution of the problem is a global optimum. For this purpose, 300 intact signals captured from the aluminum pipe *1-loc1-sml* (Figure 3.2a), at temperatures ranging from 24°C to 39°C, are used to train the SE algorithm for 10 times. The results show that the same arrivals in time trace of the signals are assigned a non-zero coefficient for all the 10 runs of the algorithm. The next question is whether the magnitude of the non-zero coefficients extracted in different runs match. At all 10 runs of the algorithm, every sample point in time is found to be assigned exactly the same value of coefficient. These results suggest stability of the proposed SE method and repeatability of the findings that will be reported in the next sections.

5.4.3 Sensitivity of the Sparse Signals to Temperature Variation (Question #3)

The results summarized in Section 5.4.1 verifies hypothesis #1, namely, the energy of guided-wave signals can be estimated using only a sparse subset of the arrivals. The next question to be answered is whether the extracted subset is less affected by EOC variations compared to a complete diffuse-field signal. This section addresses this question.

Signals captured from the intact aluminum pipe *1-loc1-sml* (Figure 3.2a) are used. The SE algorithm is trained with six different values of regularization scalar, namely, $\xi \in \{0.01, 0.05, 0.1, 0.5, 1, 10\}$. A total of 300 intact signals, recorded at temperatures ranging from 24°C and 38°C, are randomly selected for training. The rest of the signals (6,200 intact signals) are used for analyzing the sensitivity to temperature.

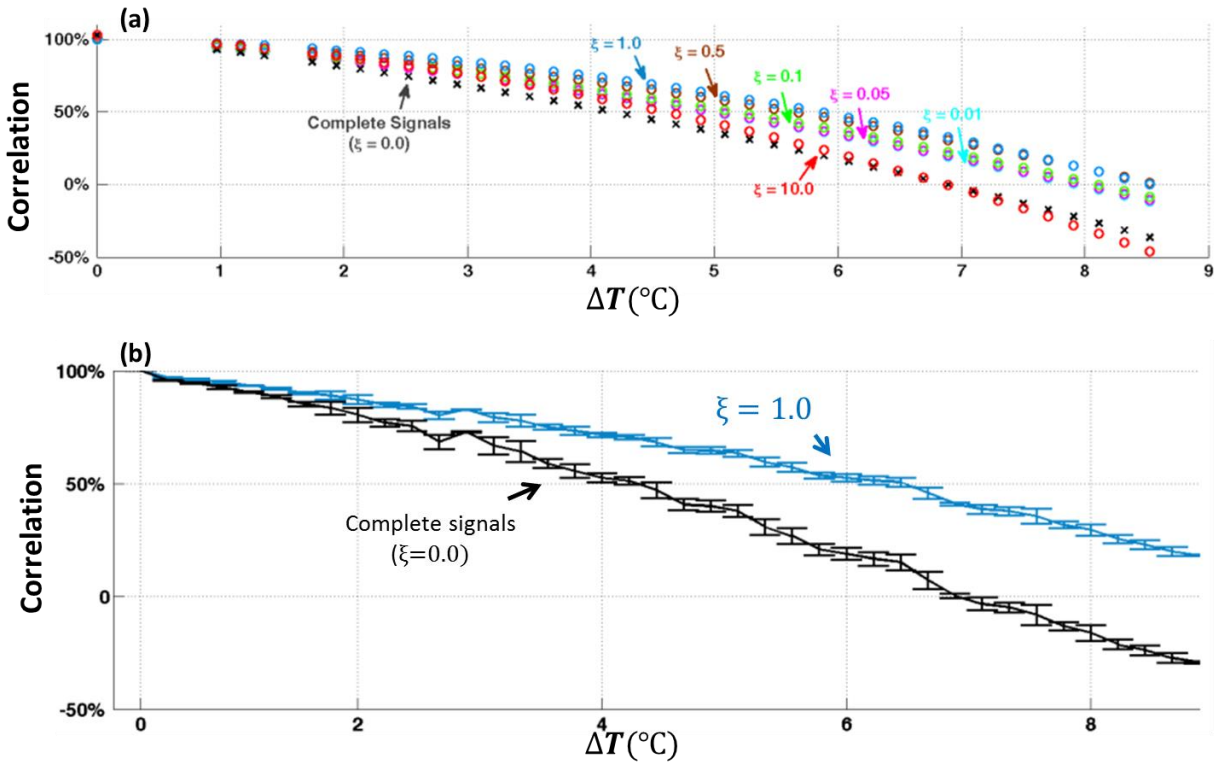


Figure 5.4: (a) Correlation between a reference complete/sparse signal ($T = 24^\circ\text{C}$) and complete/sparse test signals at different temperatures, for different values of ξ . (b) Deviation range for the correlations corresponding to $\xi = 0$ and $\xi = 1.0$. Note: Error bars represent 2σ .

Correlation between the test signals and a reference signal at temperature $T_0 = 24^\circ\text{C}$ is used as a metric to quantify the sensitivity to temperature variation. As the temperature difference (ΔT) between a test signal and the reference signal increases, the correlation between the two signals

is expected to decrease. Six \vec{X} vectors, corresponding to six different values of ξ , are used to extract the sparse test signals. That is, only the arrivals corresponding to the nonzero coefficients in \vec{X} are retained and the rest of the signal is disregarded. Average correlations of sparse signals are reported in Figure 5.4a, for ΔT s ranging from zero to approximately 9.0°C. These correlations are compared to the correlations of complete signals throughout the same range of ΔT s. For both sparse and complete signals, as the ΔT between reference signal and the test signal increases, the correlation decreases. Figure 5.4b shows that the average values of the correlations for complete signals falls beyond the standard deviation range of the correlations found for $\xi = 1.0$. As explained in previous sections of this thesis, current temperature compensation methods (*e.g.*, Clarke et al. (2009), and Konstantinidis et al. (2007)) can only improve these correlations for very small temperature differences (*i.e.*, 0.5-1°C, depending on the complexity of the structure).

However, it is observed that by imposing the sparsity (even with small regularization factor $\xi = 0.01$), the rate of this decrease is reduced significantly compared to the complete signals. As we raise the regularization scalar from 0.01 to 1.0, *i.e.*, increase the sparsity, the correlations increase for different ΔT s (*i.e.*, sensitivity of the signals to ΔT decreases). It is worth noting that $0.01 \leq \xi \leq 1.0$ corresponds to the range with close-to-zero training errors shown in Figure 5.3. These observations suggest that simplifying guided-wave signals to a sparse subset of the high-energy arrivals does improve their sensitivity to temperature variations. This addresses research question #3, and verifies hypothesis #2. Interestingly, by raising the ξ value to 10.0 the correlations drop even below the value of the correlations of the complete signals at larger values of ΔT . This observation suggests that, in addition to the tradeoff between simplicity and estimation accuracy discussed in Section 5.4.1, sensitivity to the EOCs should be considered for

selecting suitable values of ξ for optimal damage detection. An approach to effectively make such tradeoffs is summarized in Section 5.4.7.

5.4.4 Damage Detection with SE Method (Question #2)

Comparing the SE Coefficients for Intact and Damaged Signals

As discussed before, effects of damage may alter the location in time where different arrivals appear in the signal. Therefore, as stated in Hypothesis #3, using the sparse representation of intact signals for estimating the energy of damaged signals is expected to result in less accurate values than intact signals. In other words, energy estimation error (*i.e.*, ΔE in equation 5.5) is expected to increase as damage occurs. This section investigates this hypothesis through a set of laboratory experiments.

Comparing complex diffuse-field signals from intact and damaged pipes is not as simple as it may be for single-mode records. For hypothesis #3 to be valid, however, one needs to compare only the sparse subset of high-energy arrivals in these signals, rather than the complete time traces. The proposed SE algorithm (equation 5.4) is trained with a set of intact and damaged signals, separately. The goal is to compare the sparse representation of intact and damaged signals to determine whether the time-location of the extracted sparse arrivals matches. Hypothesis #3 will be rejected if there is significant overlap between the two subsets, implying that damage does not change high-energy components of the signal.

Signals captured from the intact and damaged aluminum pipe *I-locI-sml* (Figure 3.2a) are used. The temperature of both datasets ranges from 24°C to 39°C. Figure 5.5a and 5.5b show the normalized coefficients of intact and damaged signals, respectively. It is observed that only 26% of the coefficients corresponding to intact signals overlap with the coefficients of damaged signals. In other words, the arrival times of 74% of the high-energy components in the intact signals do not match with the arrival times of high-energy components in damaged signals. This mismatch can be due to two possible effects of damage: (1) change in the time location of high-energy arrivals, and/or (2) change in the energy content of the arrivals. It is also notable that the overlapping arrivals contain only about 6% of the total energy of the signals. Moreover, the correlation between intact and damaged coefficients is 16%, which further indicates the dissimilarity between the two subsets.

In order to evaluate hypothesis #3, in addition to the time location of the coefficients, I compare their corresponding magnitudes. Figure 5.5c shows the difference between the two sets of coefficients, at every point in time ($\Delta x_i = x_i^{int} - x_i^{dmg} \forall i \in \{1, \dots, n\}$). Positive values of Δx_i correspond to the arrivals with larger coefficients in the intact signals than in the damaged signals, and vice versa. The correlation of the positive side of Figure 5.5c with the intact coefficients in Figure 5.5a is 96%, while its correlation with the damaged coefficients is only 3%. Similarly, the negative side of Figure 5.5c and the damaged coefficients in Figure 5.5b are 93% correlated, while its correlation with intact coefficients is only 2%. These observations suggest that when $\Delta x_i > 0$, that is, when the coefficient of the i th arrival is larger in intact signals compared to damaged signals ($x_i^{int} > x_i^{dmg}$), x_i^{dmg} is close to zero (i.e., $x_i^{int} - x_i^{dmg} \approx x_i^{int}$). Similarly, when $\Delta x_i < 0$, that is, when $x_i^{dmg} > x_i^{int}$, x_i^{int} is close to zero (i.e., $x_i^{int} - x_i^{dmg} \approx -x_i^{dmg}$).

$x_i^{dmg} \approx x_i^{dmg}$). Therefore, not only the time location of the high-energy arrivals, but also the magnitudes of the corresponding coefficients differ substantially between intact and damaged signals.

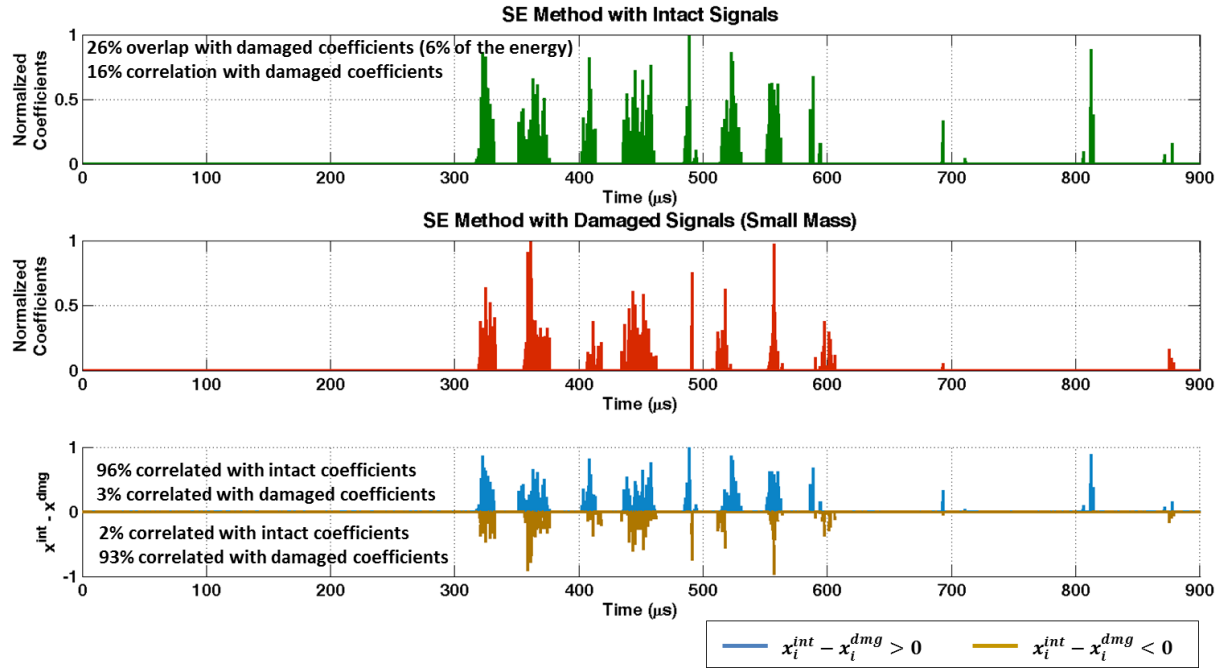


Figure 5.5: (a) The coefficients obtained by applying the SE method on intact signals, (b) the coefficients obtained by applying the SE method on damaged signals, (c) the difference between (a) and (b), i.e., $X^{int} - X^{dmg}$.

The same set of analyses is repeated with the aluminum pipe exposed to a bigger mass scatterer (i.e., *1-loc1-big* layout shown in Figure 3.2a). Results are similar to the pipe with smaller mass discussed above. 25% of the nonzero coefficients calculated for intact signals overlap with the coefficients obtained for the damaged signals. The overlapped arrivals contain only 10% of the total energy of the signals. Moreover, the correlation between the intact and damaged coefficients is 10%.

These findings imply that a sparse subset of high-energy arrivals extracted from intact signals will not be a good representative of damaged signals. Therefore, if these coefficients are used for estimating the signals' energy, the estimation errors (ΔE s) for damaged signals will be larger than the ones for intact signals. This verifies hypothesis #3, and suggests that ΔE s can be used as damage-sensitive features for the purpose of detection.

Comparing the SE and the SD Coefficients

In this section, the sparse representation of damaged signals obtained by the SE method is compared with the sparse representation of the signals obtained by the SD method (*i.e.*, supervised method discussed in Chapter 4). As discussed in Chapter 4, the coefficients obtained by the SD method correspond to the arrivals with significant damage information. On the other hand, the coefficients obtained by the SE method correspond to the arrivals that contain the majority of the signal's energy. A small overlap is expected between the SE coefficients and the SD coefficients, since one of the major effects of damage on guided-waves is energy dissipation (Long et al., 2003b).

The SE algorithm is trained with the damaged signals explained earlier in this section. Figure 5.6a shows the SE coefficients for these signals. The SD algorithm (equation 4.2) is trained with a dataset including both intact and damaged signals. Figure 5.6b shows the SD coefficients. No overlap is found between the two sets of coefficients. Same results are obtained when both algorithms are trained on the pipe with bigger mass. This suggests that damage causes significant energy dissipation in the signal so that none of the damage-sensitive arrivals has major contribution in the total energy of the signal. These results strengthen the findings discussed

earlier for verifying hypothesis #3 (Figure 5.5). That is, energy dissipation will alter the signals, thus, if the SE algorithm is trained with intact signals, energy estimation error ΔE will increase as damage occurs.

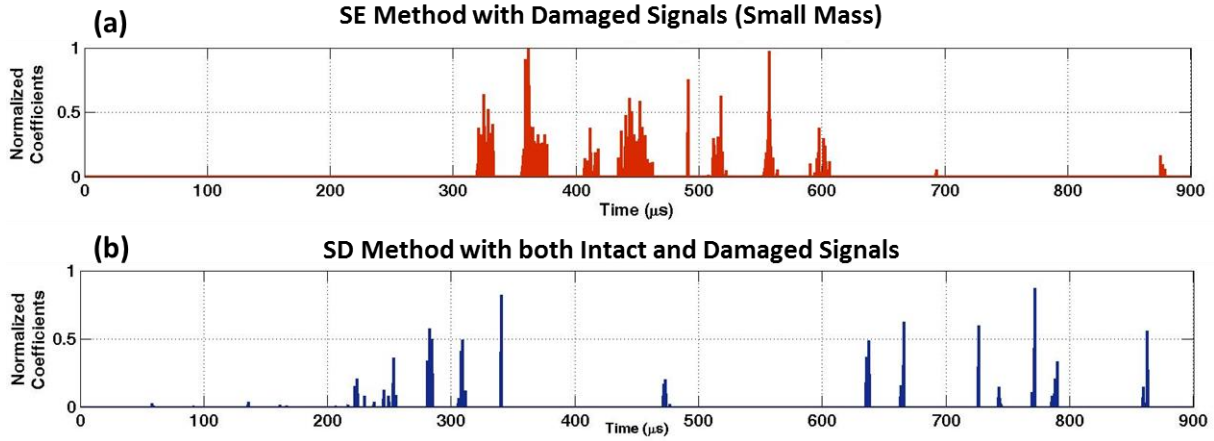


Figure 5.6: (a) The coefficients obtained by applying SE method on damaged signals, (b) the coefficients obtained by applying SD method on a training dataset including both intact and damaged signals.

5.4.5 Sensitivity of Damage-sensitive Features (ΔE s) to Temperature Variation (Question #3)

Experiments reported in Section 5.4.3 shows that sparse signals have less sensitivity to temperature variations compared to original diffuse-field signals. Later, in Section 5.4.4, it is shown that energy estimation error (ΔE) can be used as damage-sensitive features for detection purpose. However, one question that remains to be answered is the sensitivity of these features to EOC variations. This section evaluates the sensitivity of ΔE s to temperature variations.

In spite of their low sensitivity to temperature variations, time location of high-energy arrivals extracted by SE method may still vary as temperature changes. For example, let us consider a scenario in which the SE algorithm is trained with a limited range of temperature. During the monitoring stage, if the temperature of an intact signal deviates significantly from the training temperature, the sparse subset may not be a good representative of the test signal. In that case, the energy estimation error ΔE for the test signal may largely deviate from zero. This can lead to false positive errors. Therefore, the following hypothesis is evaluated in this section:

Hypothesis #4: The wider the temperature range of the training dataset, the more robust the extracted sparse subset, and thus the ΔE values, are to temperature variations.

A total of 6,200 intact signals from the aluminum pipe *1-loc1-sml* (Figure 3.2a) are used to create a number of training datasets with different ranges of temperature variation ($\delta T \in \{2^\circ\text{C}, 4^\circ\text{C}, 6^\circ\text{C}, 8^\circ\text{C}, 10^\circ\text{C}, 12^\circ\text{C}, 14^\circ\text{C}\}$). Depending on the δT range, the lower bound temperatures are between 24°C and 37°C , and the upper bound temperatures are between 26°C and 39°C . For example, to create training datasets with temperature variation range of $\delta T = 2^\circ\text{C}$, the first dataset includes all the signals between $T_1 = 24^\circ\text{C}$ and $T_2 = 26^\circ\text{C}$, the second dataset includes all signals between $T_1 = 25^\circ\text{C}$ and $T_2 = 27^\circ\text{C}$, and so on until the last dataset which includes all signals between $T_1 = 37^\circ\text{C}$ and $T_2 = 39^\circ\text{C}$. In this way, for any training δT range, different magnitudes of temperature are considered. The SE algorithm is trained with each of these datasets.

A test dataset, including 300 intact and 300 damaged observations is used. The signals in the test dataset are from a wide range of temperatures ($24^\circ\text{C} \leq T_{tst} \leq 39^\circ\text{C}$). Using the coefficients

obtained from each of the training datasets explained above, energy estimation error (ΔE) is calculated for the test signals. Correlations of these ΔE s with temperature of the test signals are calculated. Figure 5.7 reports these correlations with respect to training temperature range. Figure 5.7a is the correlation of ΔE s for intact test signals, and Figure 5.7b is the correlation of ΔE s for damaged test signals. It can be seen that, by increasing the training temperature range, the correlation between the test ΔE s and test temperatures drops, which verifies hypothesis #4. That is, the wider the range of training temperature, the less sensitive to temperature variations is the extracted sparse subset. The SE algorithm enforces solutions that represent the majority of training signals' energy and are robust to the factors varying among the signals (in this case, temperature effects). Therefore, the wider the training temperature, the more robust are the extracted coefficients to temperature effects. It is also notable that the variance in the correlation values diminishes as the training temperature range increases.

As training temperature range increases, correlation of the ΔE s of intact signals drops more than the correlation of ΔE s of damaged signals. This difference is expected, because the sparse subset extracted from intact training signals do not correspond to the high-energy, and thus more robust, arrivals in the damaged signals. Therefore, as can also be inferred from the figure, independent of training temperature range δT , when damage occurs, in addition to the increase in energy estimation error ΔE (as discussed in Section 5.4.4), sensitivity of ΔE s to temperature variations may also increase

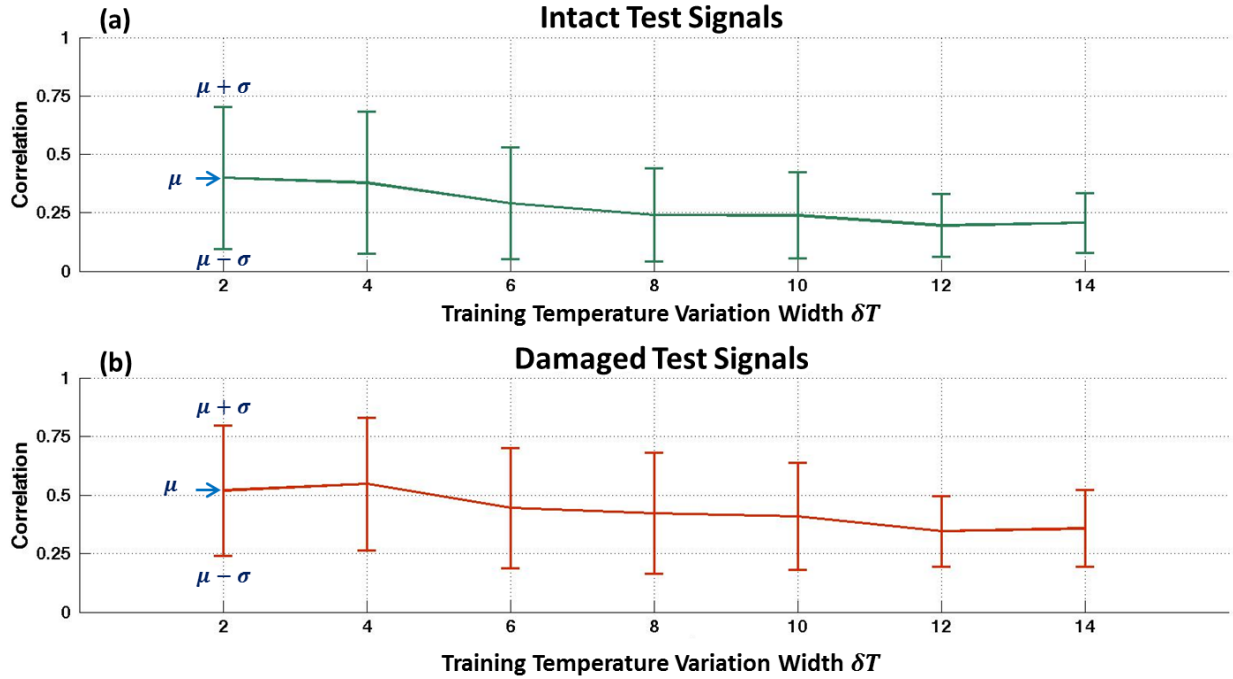


Figure 5.7: Correlation of energy estimation errors (ΔE_s) of (a) intact test signals and (b) damaged test signals, for different widths of training temperature range (δT). Note: δT is the difference between the maximum and minimum temperatures within different training datasets. μ is the mean of the correlations at each δT , and σ is the standard deviation of the correlation values at each δT .

5.4.6 Physical Intuition about the SE Coefficients

In order to better understand the findings reported in this chapter, a set of experiments are conducted on the aluminum pipe *2-loc2-sml* (Figure 3.2d). In this layout, the two transducers are located as close as 1/3 of the length of the pipe, and the mass scatterer (to simulate damage) is located on the opposite side of the pipe. In this case, the length of the shortest damaged path, as well as end-reflection paths, is much longer than the length of the shortest undamaged path. This will result in better separation of the first arrivals from damaged and undamaged sections of the

pipe. Therefore, using this layout, one can better understand the physical nature of what is extracted by the SE algorithm.

A total of 300 intact signals, and 300 damaged signals, with temperatures varying between 24°C and 38°C are used. The SE algorithm is trained using the intact signals, as well as the damaged signals, separately. In addition, both the intact and damaged signals are used to train the SD algorithm (*i.e.*, supervised method discussed in Chapter 4). Figure 5.8a and 5.8b show the normalized coefficients of the SE method when using intact and damaged signals, respectively. Figure 5.8c is the normalized coefficients obtained by supervised SD method. The expected time-location for first arrivals from the undamaged path and the first arrivals from the damage are marked in the figure.

It is found that 42% of the coefficients in Figure 5.8a (*i.e.*, SE method with intact signals) overlap with the coefficients in Figure 5.8b (*i.e.*, SE method with damaged signals). This overlap rate is about 17% larger than what was found in Section 5.4.4 with *1-loc1-sml* and *1-loc1-big* pipe layouts. This can be due to the very short paths in the layout used in this section. The shorter travel path, and thus shorter travel time, can decrease the attenuation in *2-loc2-sml* layout compared to the layouts used in Section 5.4.4 (considering that the pipe condition is the same for all pipe layouts). Therefore, part of the difference in the intact and damaged coefficients observed in Section 5.4.4 can be attributed to the larger attenuations of the waves traveling through longer damaged paths as compared to undamaged paths.

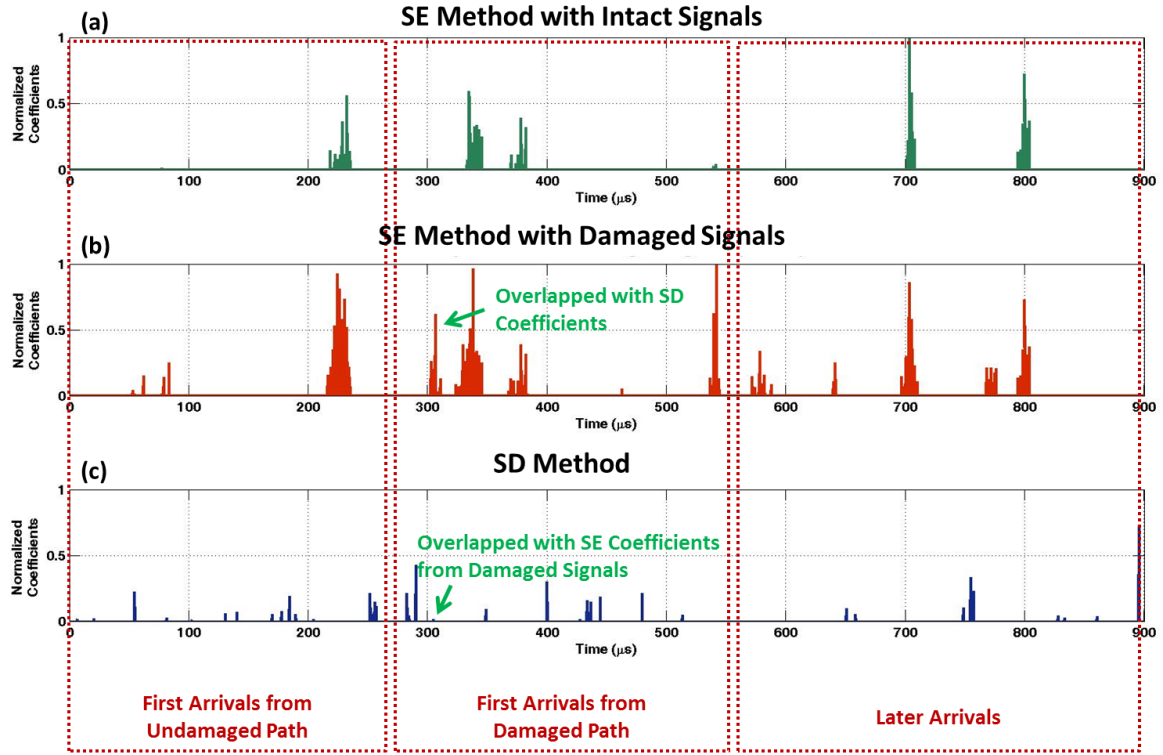


Figure 5.8: Normalized coefficients obtained by (a) SE method using intact signals, (b) SE method using damaged signals, and (c) SD method using both intact and damaged signals.

Note: the aluminum pipe layout *2-loc2-sml* shown in Figure 3.2d is used for these experiments.

Table 5.1: The breakdown of the total energy, and the overlap between the coefficients extracted by SE method, using intact and damaged signals separately. **Note:** The aluminum pipe layout *2-loc2-sml* shown in Figure 3.2d is used for these analyses.

	Energy (% of total)		Overlap between Intact and Damaged SE Coefficients
	Intact Signals	Damaged Signals	
First arrivals from undamaged path	28%	27%	10%
First arrivals (from damaged path)	38%	39%	25%
Later arrivals	34%	34%	7%

The overlapped arrivals contain only 4.4% of the total energy of the intact signals, and 4.2% of the total energy of the damaged signals. Table 5.1 summarizes the energy breakdown between the three arrival zones shown in Figure 5.8. It can be seen that the second set of arrivals contains slightly larger portion of energy, compared to the first and later arrivals. Moreover, the majority of the overlapped coefficients between intact and damaged SE coefficients are among the second set of arrivals (*i.e.*, 25% of the total 42% overlap). The question to be answered here is whether the overlapping high-energy arrivals in the second zone are from damaged sections of the pipe. Due to damage effects, it is expected that the majority of these arrivals are not reflected from damage, but rather are the second arrivals from the undamaged sections of the pipe.

One way to examine whether the overlapping arrivals in the second zone include the reflections from damage is to compare these coefficients from the ones extracted by the SD method. By nature, the coefficients obtained by SD method correspond to the arrivals with significant damage information. It is found that only 0.4% of the SD coefficients (*i.e.*, only one coefficient) in Figure 5.8c overlap with the SE coefficients obtained from damaged signals (Figure 5.8b). This coefficient is marked in Figure 5.8b and 5.8c. Moreover, no overlap is found between SD coefficients and the SE coefficients from intact signals. These findings strengthen the arguments presented in Section 5.4.4 regarding the energy dissipation effects of damage. That is, the arrivals containing significant damage information are not the ones with high energy content.

Table 5.2 summarizes the energy breakdown of the sparse signals. It can be seen that the overall amount of energy covered by the second set of arrivals is ~10% lower for damaged signals compared to intact signals (*i.e.*, 60% versus 71%). On the other hand, for damaged signals, the

contribution of the first set of arrivals in the total energy is almost two times larger than the intact signals (*i.e.*, 30% versus 16%). These changes can be attributed to the drop in the energy content of the second and later sets of arrivals due to existence of damage. Therefore, for damaged signals, the share of the first arrivals in the signals' total energy increases.

Table 5.2: The breakdown of the total energy of the sparse signals, and the overlap between the coefficients extracted by SE method, using intact and damaged signals separately. Note: The aluminum pipe layout *2-loc2-sml* shown in Figure 3.2d is used for these analyses.

	Energy of sparse signals (% of total)		Overlap between Intact and Damaged SE Coefficients
	Intact Signals	Damaged Signals	
First arrivals from undamaged path	16%	30%	10%
First arrivals (from damaged path)	71%	60%	25%
Later arrivals	13%	10%	7%

Minimal overlap between the SD and SE coefficients suggests that the SE coefficients of damaged signals in the second and third zones (Figure 5.8b) can correspond to either of the following: (1) high-energy arrivals reflected from damage that are less affected by damage interference, thus are not associated with large SD coefficients (such as the one marked in Figure 5.8c), (2) high-energy arrivals from undamaged sections of the pipe, which can be different from the ones captured from intact pipe due to superposition of the reflections from damage.

Another noteworthy point is that only 10% of the arrivals in the first zone of the intact signals overlap with those of damaged signals. In the damaged pipe, these arrivals are captured before the reflections from damage are received. Therefore, one may expect large overlap between the intact and damaged coefficients in this section of the signals. However, damage changes the physical properties of the pipe (*e.g.*, stiffness, young modulus, *etc.*), which can affect the time-location of the high-energy arrivals. To elaborate more on this fact, let us consider the SD coefficients associated with this part of the signal (Figure 5.8c). It can be seen that a number of these arrivals are assigned non-zero SD coefficients. This implies that these arrivals contain significant damage information. Since none of these arrivals includes reflections from damage, it can be concluded that the damage does affect the physical properties of the pipe, and hence, the arrival time of the wave modes traveling through undamaged sections of the pipe. This justifies the low overlap between intact and damaged SE coefficients in the first zone.

5.4.7 Selecting Regularization Scalar (ξ) for SE Method (Question #1)

As it was shown in Section 5.4.1, by increasing the sparsity constraint (*i.e.*, increasing regularization scalar ξ in equation 5.4), sparser solutions are obtained, at the cost of a decrease in energy estimation accuracy. Therefore, for selecting the ξ value, one tradeoff to be considered is between sparsity and the training error. On the other hand, in Section 5.4.3, I showed that increasing the sparsity could decrease the sensitivity of the sparse signals to temperature variations. However, too sparse solutions can be more sensitive to temperature variations than complete signals. Therefore, the objective is to select a ξ value that maximizes sparsity, while achieving minimal training error and sensitivity to EOC variations.

Intact signals from the aluminum pipe *1-loc1-sml* (Figure 3.2a) are used to train the SE algorithm with different ξ values, ranging from 0 to 950. The temperatures of the training signals vary between 24°C and 39°C. For simplicity, I used bigger increments between the ξ values for larger magnitudes (see Figure 5.9). The coefficients corresponding to each ξ are used to calculate the energy estimation error for every signal in the training dataset ($\Delta E_j \forall j \in \{1, \dots, m_{tr}\}$), as given in equation 5.5. The correlation of the ΔE_j values to temperature of the training signals is used as a metric to reflect their sensitivity to temperature variation.

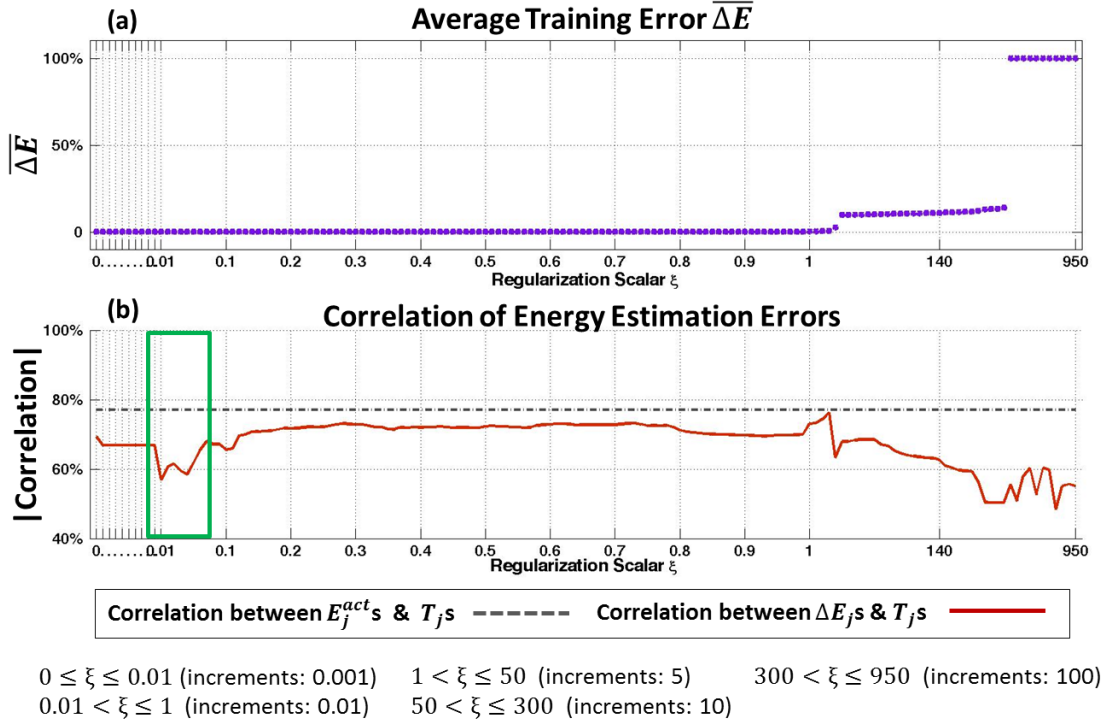


Figure 5.9: (a) Average training error at different values of regularization scalar ξ (see equation 5.6), (b) Correlation between the energy estimation errors ($\Delta E_j \forall j \in \{1, \dots, m_{tr}\}$, see equation 5.5) and the temperatures of training signals (T_j) at different values of regularization scalar ξ . E_j^{act} : energy of the complete training signals.

Figure 5.9b shows ΔE_j s and temperature correlations for different values of ξ . It can be seen that by increasing the sparsity, the sensitivity of the ΔE values to temperature decreases. The correlations can reach values up to 20% less than the correlation of the complete signals' energy (shown with a dashed line). By further increasing the sparsity, however, the correlation rises back. These findings are in agreement with the results reported in Section 5.4.3. That is, sparsity decreases the sensitivity of the signals to temperature variations. However, too sparse solutions can be more sensitive than complete signals. It is notable that, at large values of ξ , the correlation drops again to its minimum value. These larger values of ξ correspond to large training errors. This trend is expected since too sparse solutions include very few arrivals (usually one arrival with very small coefficients), which cannot be correlated to temperature. These solutions would be rejected due to large training error. Based on this approach, for the training signals used in this figure, an optimum value of ξ would be between 0.01 and 0.05 (marked with a green box in Figure 5.9b).

Physical Intuition:

In order for better understanding of the pattern observed in Figure 5.9, let us use the aluminum pipe layout *2-loc2-sml* (Figure 3.2d). As mentioned before, the first and later arrivals are more distinguishable in this layout, facilitating the interpretation of the findings. Figure 5.10 shows the average training error ($\overline{\Delta E}$), as well as the correlation between the temperature and energy estimation errors (ΔE_j). Based on the approach discussed above, for this pipe layout, suitable ξ values would be between 0.25 and 0.33.

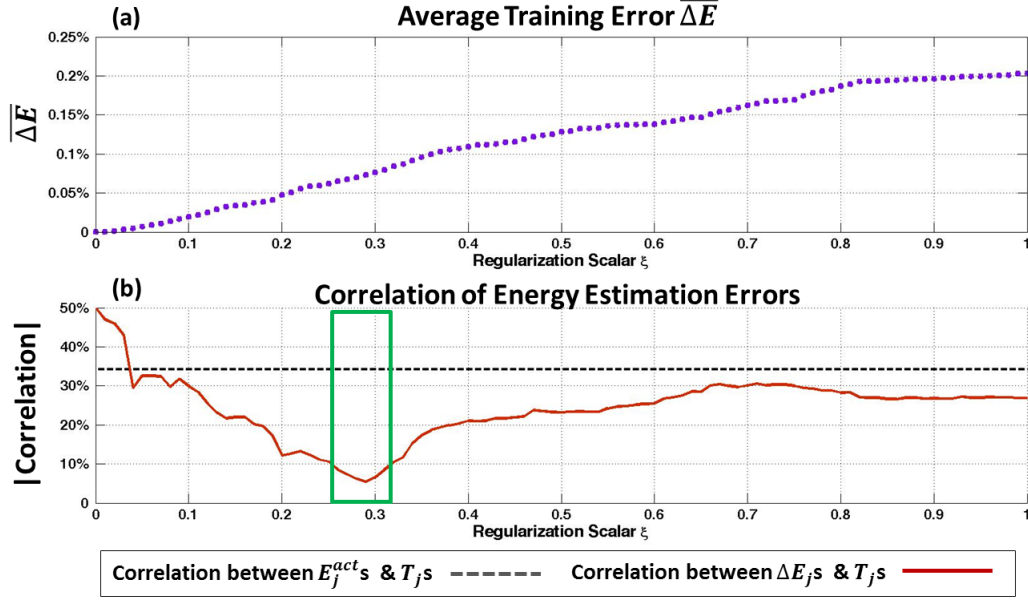


Figure 5.10: (a) Average training error at different values of regularization scalar ξ (see equation 5.6), (b) Correlation between the energy estimation errors ($\Delta E_j \forall j \in \{1, \dots, m_{tr}\}$, see equation 5.5) and the temperatures of training signals (T_j) at different values of regularization scalar ξ . E_j^{act} : energy of the complete training signals.

Similar to Section 5.4.6, the coefficients are studied with regard to three sets of arrivals, namely first arrivals from undamaged paths, second set of arrivals (if the pipe is damaged, would include first arrivals from damaged path), and the later arrivals. Figure 5.11 shows the change in the percentage of non-zero coefficients as ξ value is increased, at each of the three arrival zones. It can be seen that as sparsity constraint is increased, the majority of the first and later arrivals are set to zero much faster than the second arrivals. In the ξ range in which the optimal temperature correlation is observed (*i.e.*, $0.25 \leq \xi \leq 0.33$, marked with green box in Figures 5.10 and 5.11), the majority of the coefficients in the first and later arrivals are set to zero. However, it is notable that the few nonzero coefficients in these zones remain almost unchanged by further increasing the ξ . Among the three zones, the majority of the non-zero coefficients fall in the second set of arrivals. This is expected since according to Table 5.2, 71% of the energy of the intact signals is

represented by the second set of arrivals. Unlike the first and later coefficients, in the second zone, the drop in the percentage of non-zero coefficients continues as ξ increases. Therefore, as the ξ value is increased beyond the optimum range, the increase in the training error, and in the temperature sensitivity, is mainly because of losing the high-energy arrivals in the second zone.

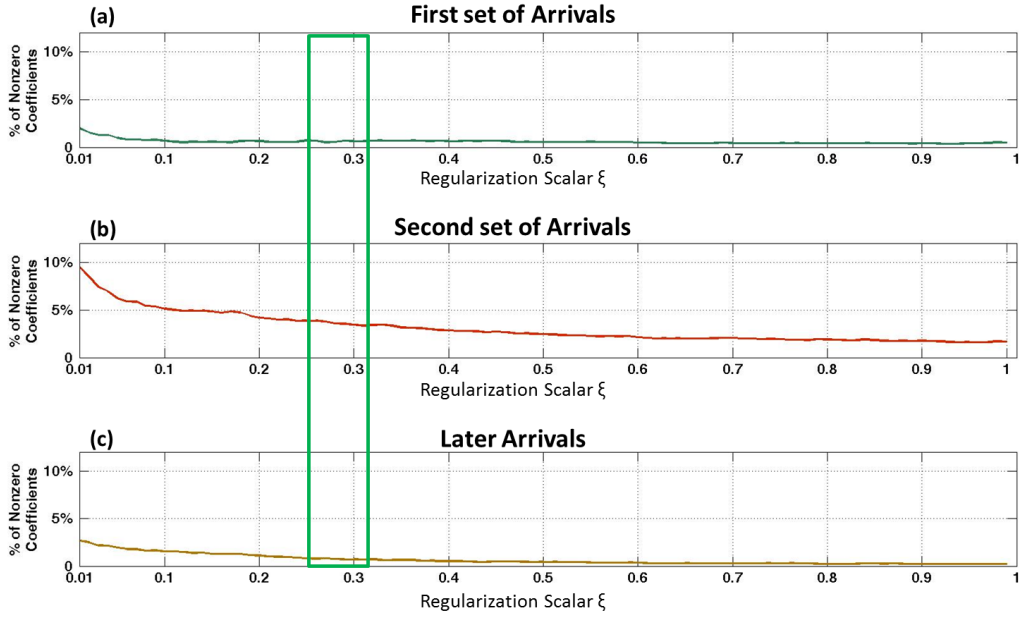


Figure 5.11: Change in the percentage of nonzero coefficients for different levels of sparsity (*i.e.*, different values of regularization scalar ξ) corresponding to (a) first set of arrivals, (b) second set of arrivals, and (c) later arrivals (see Figure 5.8 for clarification on first, second and later set of arrivals).

5.5 Validation of the SE Method

In this section, signals captured from laboratory and field testbeds are used to validate the discriminatory power of the damage-sensitive features proposed in this chapter.

5.5.1 Methods for Evaluating the Classification and Detection Power of ΔE s

At this point, it is worth emphasizing that, similar to SD method discussed in Chapter 4, the SE method is not a detection algorithm, but a feature extraction approach for damage detection. That is, the contribution of this chapter is in proposing an unsupervised feature extraction method for continuous damage detection of pipelines under varying EOCs. A number of detection methods may be used to evaluate discriminatory power of the features. However, unlike SD method, the SE method proposed in this chapter is an unsupervised approach. Therefore, to fulfil the objective of the SE method, an unsupervised detection approach should be applied for online damage detection using the extracted damage-sensitive features (ΔE s). Development of an optimal online detection algorithm is beyond the scope of this thesis. Nonetheless, in order to evaluate the discriminatory power of damage-sensitive features, to better understand the distribution of features in both classes (*i.e.*, intact versus damaged), and to prove the concept of online damage detection with the SE features, three sets of detection/classification approaches are used for the experiments in this section that will be discussed in the following sub-sections.

A Heuristic Unsupervised Classification based on K-means

K-means is a simple unsupervised partitioning method in which each observation is assigned to one cluster. The performance of the K-means in separating the two classes will provide insight about the separation of the ΔE values in two populations. First, K-means is used to cluster the test observations based on the corresponding ΔE s. Some heuristic process is then applied to predict the class labels. If more than 50% of the observations in a cluster are from the same class (*e.g.*, intact), the cluster is considered to be representative of that particular class. Then,

accuracy, FPR, and FNR can be calculated based on the predicted labels. Good detection performance with K-means implies concentrated values of ΔE s for a particular class, which suggests that occurrence of damage may be detectable with some simple outlier or threshold based approach. Expectedly, K-means will perform well if increase in the average ΔE values with occurrence of damage is larger than their variance in either of the classes.

A Heuristic Supervised Classification based on Support Vector Machin (SVM)

Support vector machines (SVM) are supervised learning methods that can be used for constructing a hyperplane in a high-dimensional space so that its distance to the nearest training feature of any class is maximized (Cortes and Vapnik, 1995). In this section, SVM with Gaussian Radial Basis Function (RBF) kernel is used in order to identify cases where a simple linear discriminator, like K-means, may not be sufficient for optimal separation of the two classes. As mentioned before, the proposed SE method is an unsupervised approach. However, to apply SVM, our prior knowledge about the class labels needs to be incorporated. It is worth re-emphasizing that the SVM-based classification method applied here is not proposed as a detection approach, but is used to validate the discriminatory power of ΔE s and to understand the nature of the separation between the ΔE s in two classes. For each set of experiments that will be discussed in this section, 100 signals from the intact pipe and 100 signals from the damaged pipe are used to train the SVM algorithm. That is, these signals are projected on the extracted sparse X vector, ΔE s are calculated, and used to train SVM. The trained SVM model is used to predict class labels of a different set of test signals, based on their corresponding ΔE values.

An Online Detection Method based on Kullback-Leibler Divergence (KL)

Although the two methods described above evaluate the separation between the two populations, such approaches cannot be applied for online detection of damage. First, both of them are based on our prior knowledge regarding the class labels. Second, using pure classification methods, we lose the information regarding the order in which the features are observed, as well as the proximity of the observations. For example, an intact observation is not likely to happen after an observation from the damaged pipe. Similarly, if the i th observation is from an intact pipe, it is more likely than not that the $(i+1)$ th observation is also from an intact pipe. Finally yet importantly, when damage occurs, the total population of observed ΔE s may not be sufficiently separable from all the ΔE s corresponding to intact pipe. However, this does not necessarily mean that the instance of occurrence of damage cannot be detected. As I discussed in previous sections of this chapter, when damage occurs, not only the ΔE values are expected to increase, but also the sensitivity of these values to EOC variations may increase compared to the intact observations. This implies that the distribution of the ΔE values is expected to change as damage occurs (see Figure 5.13 for an example).

KL divergence method is a measure of the difference between two probability distributions (P & Q) (Kullback and Leibler, 1951). In other words, KL measures the amount of the information that would be lost if one distribution (Q) is used to approximate the other (P). For discrete probability distributions, KL divergence is defined as below:

$$D_{KL}(P||Q) = \sum_x P(x) \ln \frac{P(x)}{Q(x)} \quad (5.8)$$

In this section, I use this method to detect occurrence of damage. After training the SE algorithm, the ΔE values are obtained for test signals. The first 100 ΔE s from the test intact signals are used to obtain a reference distribution (P). Therefore, this approach assumes that the pipe is in intact state when testing starts. Test distributions (Qs) are obtained by a moving window of width 50 ΔE s as new signals are observed. The KL measures are calculated for all the moving windows. Occurrence of damage is detected when KL measure deviates from its previous value by more than two times the standard deviation of all the previous measures up until the last window. Detection performance will be presented in terms of delay in detection (DiD), and false negative rate (FNR).

5.5.2 Damage Detection with Varying Training and Test Temperatures

The temperature range that is covered by the training dataset, as well as the temperature variations among the test signals can both affect the discriminatory power of the damage-sensitive features (*i.e.*, energy estimation errors, ΔE s). In this section, detection performance of the SE method is validated for different scenarios regarding these factors.

The aluminum pipe layout *I-locI-sml* (Figure 3.2a) is used. A total of 6,500 signals from the intact pipe and 6,300 signals from the pipe with structural abnormality are captured, at temperatures ranging from 24°C to 38°C. Seven training subsets of intact signals, each including 300 observations, are selected so that their temperature range increases from 2°C to 14°C, with 2-degree increments. In other word, the difference between maximum and minimum temperature in the first subset is 2°C, and in the seventh subset is 14°C. A test dataset of 200 intact and 200

damaged signals, with temperatures ranging from 24°C to 39°C is used. The SE method is trained with each of the training datasets. Using the coefficients corresponding to each training run, the ΔE values are obtained for the test observations.

Table 5.3: Average detection statistics for 10-fold cross validation with the ΔE values calculated with the coefficients corresponding to different ranges of training temperature.

Note: The temperatures of test signals vary between 24°C and 39°C.

	K-means			SVM			KL	
	Accuracy	FPR	FNR	Accuracy	FPR	FNR	DiD	FNR
$24^{\circ}\text{C} \leq T_{tr} \leq 26^{\circ}\text{C}$	92.5%	8.5%	6.5%	94.0%	5.0%	7.0%	17	8.5%
$24^{\circ}\text{C} \leq T_{tr} \leq 28^{\circ}\text{C}$	93.0%	8.5%	5.5%	94.5%	2.0%	9.0%	18	9.0%
$24^{\circ}\text{C} \leq T_{tr} \leq 30^{\circ}\text{C}$	97.5%	1.5%	3.5%	99.0%	1.0%	1.0%	5	2.5%
$24^{\circ}\text{C} \leq T_{tr} \leq 32^{\circ}\text{C}$	97.3%	0.0%	5.5%	97.0%	0.0%	6.0%	1	0.5%
$24^{\circ}\text{C} \leq T_{tr} \leq 34^{\circ}\text{C}$	97.8%	0.0%	4.5%	98.0%	0.0%	4.0%	1	0.5%
$24^{\circ}\text{C} \leq T_{tr} \leq 36^{\circ}\text{C}$	97.0%	0.0%	6.0%	99.5%	0.0%	1.0%	1	0.5%
$24^{\circ}\text{C} \leq T_{tr} \leq 38^{\circ}\text{C}$	97.0%	0.0%	6.0%	100.0%	0.0%	0.0%	1	0.5%

Table 5.3 summarizes the average classification statistics for 10-fold cross validation, for K-means and SVM based approaches, as well as online detection statistics for KL based approach. The K-means and SVM results suggest that ΔE values of intact and damaged signals can be separated satisfactorily, even with limited ranges of training temperature. The separation, however, improves as wider ranges of temperature are used for training of the SE algorithm. SVM performs slightly better than the K-means, particularly with wider ranges of training temperature. In terms of online detection, the KL based approach implemented here is able to

detect occurrence of damage in all cases, however, the detection performance improves considerably with wider ranges of training temperature.

5.5.3 Detection of a Small Crack and Mass Loss

As it is discussed in previous sections, detection performance of the SE method depends on the extent of the change caused by damage on high-energy arrivals of the signal. Therefore, it is important to validate the method for detection of small defects whose effects may not be significant compared to those of EOCs. In this section, the proposed SE method is used for detection of actual damages of different type and size. A small crack is simulated by a saw-cut on the surface of a steel pipe. Later, a small mass loss is introduced to the pipe, in a different location than the crack. The objective is to validate the proposed SE method for detection of small damages, as well as detection of occurrence of consecutive damages, under temperature variations.

For these experiments, a Schedule-40 steel pipe shown in Figure 3.3 is used. Before introducing the crack, ultrasonic pitch-catch records are obtained from the intact pipe, at different temperatures ranging from 25°C to 27°C. These signals are used for training the SE algorithm. Later, an oblique cut of 2.5 *cm* long, extended in both circumferential and longitudinal directions of the pipe (see Figure 3.3), is imposed at the middle of the length of the pipe. The maximum thickness along the length of the crack is approximately 0.34 *mm* (10% of the pipe thickness), and the maximum depth is approximately 0.7 *mm* (20% of the pipe thickness). The second damage is a small mass loss (Figure 3.3) with 5 *mm* diameter and maximum depth of 0.25 *mm*

(~8% of the pipe thickness). For testing, signals from the intact pipe and the pipe with crack, and with both crack and mass loss are used at $24^{\circ}\text{C} \leq T_{tst} \leq 32^{\circ}\text{C}$.

Similar to the previous section, a heuristic approach based on k-means clustering, as well as a supervised binary classification with SVM is used to evaluate the separation among the ΔE s calculated for the test signals at different states of the pipe. Moreover, the KL based detection approach discussed at the beginning of this section is applied to detect occurrence of each of the consecutive damages.

Table 5.4: Average detection statistics for 10-fold cross validation with the ΔE values calculated for the intact steel pipe, the pipe with a small crack, and the pipe with both crack and small mass loss.

State of the pipe	K-means			SVM			KL	
	Accuracy	FPR	FNR	Accuracy	FPR	FNR	DiD	FNR
Intact & Crack	94.5%	1.2%	12.4%	95.0%	6.7%	1.3%	11	4.4%
Crack & Mass Loss	98.7%	0.0%	3.9%	100.0%	0.0%	0.0%	1	0.8%

Table 5.4 reports the statistics for detection of the crack, as well as the mass loss. As can be seen in this table, both classification methods are successful in separating the ΔE s corresponding to two different states of the pipe. Expectedly, this distinction becomes more significant when the second damage is introduced. The KL based approach is also successful in detecting the occurrence of both damages. However, there is a delay of eleven observations in detection of the

first damage. Figure 5.12 is a visual depiction of the calculated ΔE s at different states of the pipe, along with the temperature readings corresponding to each test observation.

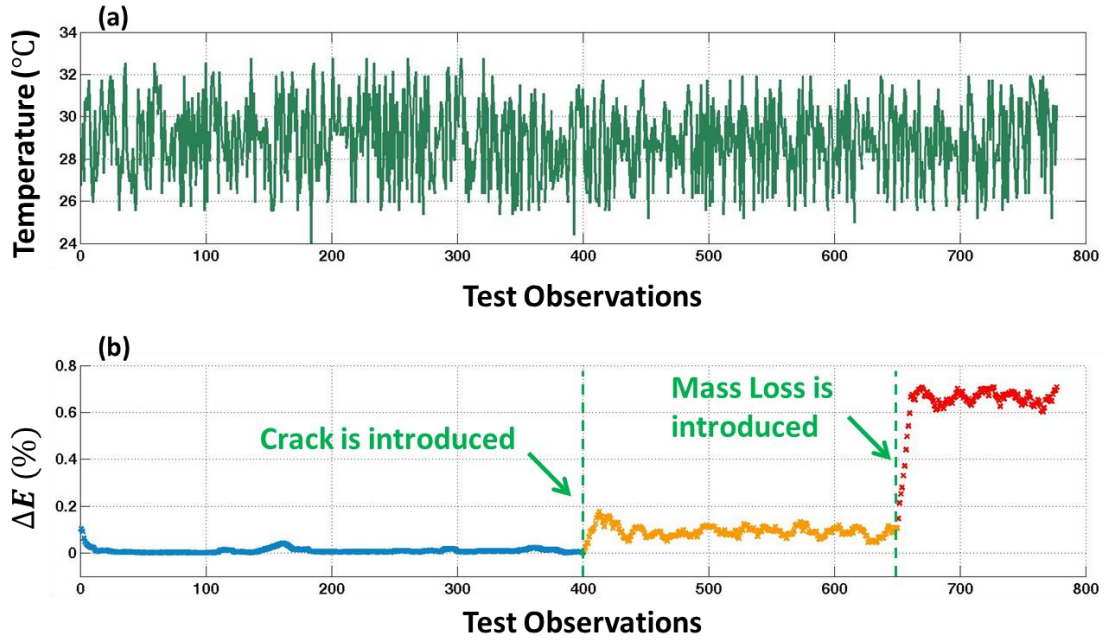


Figure 5.12: Online detection of a 2.5 cm long oblique crack (with maximum thickness of 0.34 mm and maximum depth of 0.7 mm), and a subsequent small mass loss (with maximum depth of 0.25 mm), in a Schedule-40 steel pipe.

5.5.4 Online Damage Detection under Multiple EOCs

For the experimental analyses and validations so far, temperature is the only varying EOC. In this section, the performance of the SE method for damage detection under multiple varying EOCs is validated. Signals from a hot-water piping system (Liu et al., 2012b), operating under mechanically and electrically noisy environment, with varying temperature, fluid flow rate, and

inner pressure, are used for the purpose of this section. For more details on the specifications of the piping system and the data, please refer to Section 3.3.

Table 5.5: Average detection statistics for 10-fold cross validation with the ΔE values calculated for a hot-water supply piping system, with and without a structural abnormality.

	K-means			SVM			KL	
	Accuracy	FPR	FNR	Accuracy	FPR	FNR	DiD	FNR
Winter Day #1	98.0%	0.0%	3.9%	99.1%	0.0%	1.9%	3	2.0%
Winter Day #2	98.5%	1.5%	1.5%	99.3%	0.4%	0.9%	3	0.9%
Summer Day #1	92.2%	1.5%	14.0%	94.5%	4.2%	6.7%	1	0.4%
Summer Day #2	54.3%	0.0%	91.0%	76.8%	14.2%	32.0%	1	0.8%

Table 5.5 summarizes the classification and detection statistics for two winter days and two summer days. For winter day #1 and #2, and for summer day #1, both classification methods can sufficiently separate the observations in two classes. In summer day #2, however, the features of the two classes are not as distinguishable as the ones in the other days. Nevertheless, as explained before, this does not necessarily mean that the occurrence of damage cannot be detected, since online damage detection is not a pure classification problem. As can be seen in the last two columns of Table 5.5, KL based approach can detect occurrence of damage with good precision in all of the four days. Figure 5.13 shows an example case where perfect classification of the two classes with K-means and SVM based approaches is not possible, but occurrence of damage can be detected successfully with KL based approach, because of the significant change in the distribution of the ΔE s after occurrence of damage. In this particular

case, after introducing the damage, the ΔE values become very sensitive to periodic variation of temperature and flow rate.

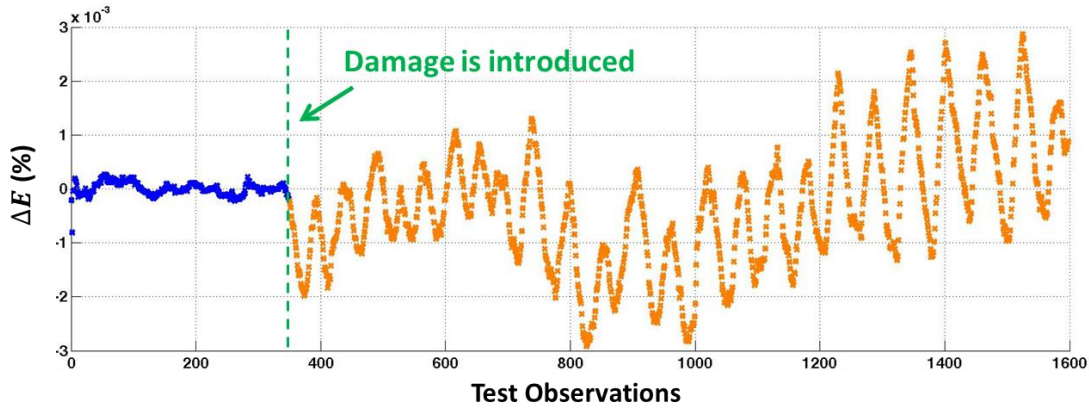


Figure 5.13: An example case from field data where perfect classification of the ΔE s with K-means and SVM based approaches is not be possible, but occurrence of damage can be detected using KL based approach.

5.5.5 Effects of Including Damaged Signals in the Training Dataset

The proposed SE method assumes that the majority of the training dataset is composed of signals from intact state of the pipe. It is important to define the majority in this context, and to understand the behavior of the method at different levels of violation of this assumption.

Signals from the aluminum pipe *1-loc1-sml* (Figure 3.2a) are used. A total of 300 intact signals are randomly selected for training the SE algorithm. The temperature of these signals varies between 24°C and 39°C. At different iterations of SE algorithm, a portion of the training intact signals is replaced with randomly selected damaged signals. The ratio of the damaged signals to the total number of training signals is increased in each iteration, from 0.3% (only one damaged

signal) to 99% (298 damaged signals). A fixed set of test signals, including 200 intact and 200 damaged observations, are used for evaluating the SE coefficients at each iteration.

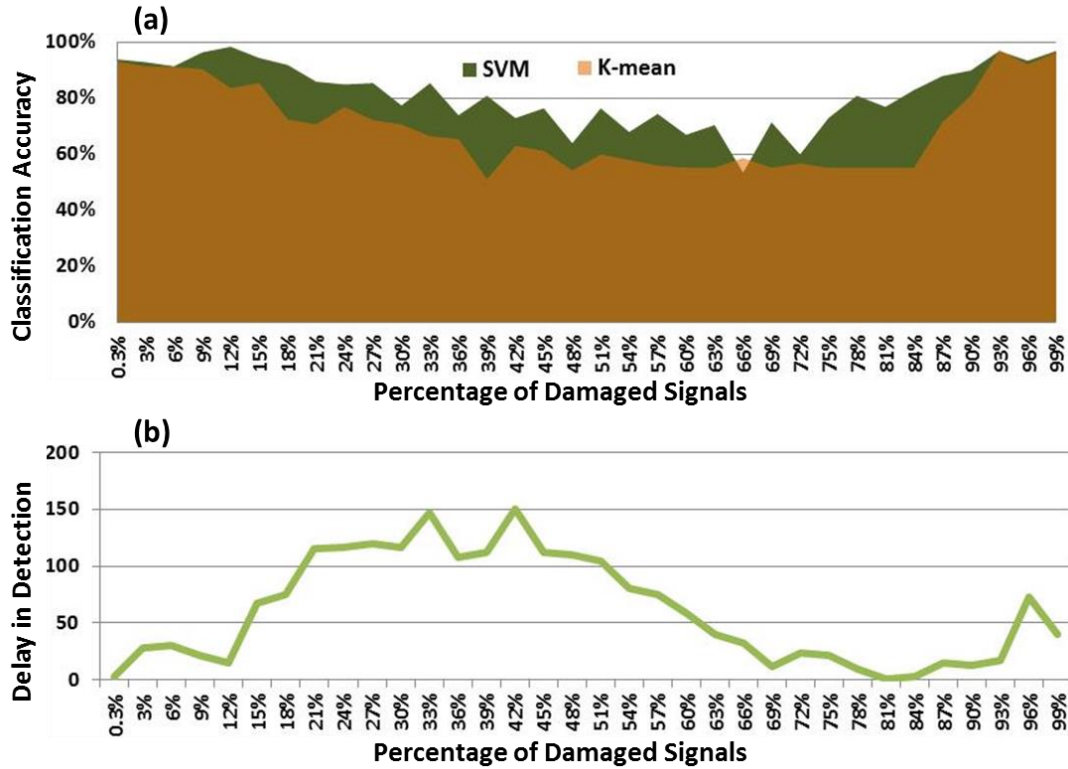


Figure 5.14: (a) Classification accuracies, and (b) delays in detection of occurrence of damage, corresponding to different ratios of damaged signals in the training dataset.

Figure 5.14a shows the classification accuracies of K-means and SVM based methods at each iteration, and Figure 5.14b shows the delay in detection of the occurrence of damage. Expectedly, as more damaged signals are included in the training dataset, the performance of the SE method in separating intact and damaged observations decreases. The classification accuracy of the SVM remains above 80% for damaged signal ratio of up to about 30%. Detection delay, however, is more sensitive to the increase in the ratio of damaged signals. Considering the

definition of KL divergence method, this behavior is expected. As the damaged signals are added to the training dataset, the extracted coefficients become more representative of damaged signals. Therefore, the distributions of ΔE s for intact and damaged test signals become more similar to each other. It is also notable that further increasing the ratio of damaged signals, the classification accuracies of both methods rise back to their maximum range. In these cases, the calculated coefficients are mostly representative of damaged signals. Therefore, energy estimation error (ΔE) of intact test signals are expected to be larger than the ones for damaged test signals. Similar to the SE method, the ΔE values can be the discriminatory feature for the two classes.

Chapter 6

Conclusion

6.1 Summary of the Work

The work presented in this dissertation addresses some of the open challenges in real-world application of guided-waves for pipeline damage detection. In Chapter 1, these challenges are discussed under three headings: (a) Multiple modes, (b) Multi-path reflections, and (c) Sensitivity to environmental and operational conditions (EOCs). The objective is to address these challenges through extracting simplified representations of guided-wave signals in which the effects of EOCs are suppressed by the effects of damage. In Chapter 4 and Chapter 5, I proposed two frameworks for online damage detection of pipelines operating under varying EOCs.

The first framework (Chapter 4) is based on a supervised method for extracting a linear sparse discriminant in the time domain of the signals (*i.e.*, Sparse Discriminant (SD) method). The idea is that a sparse subset of the arrivals will include enough information for optimal damage detection. Therefore, extracting only the arrivals with significant sensitivity to damage, both

detection performance and the sensitivity to EOC variations will improve compared the complete signals. I validated this statement through a set of laboratory and field experiments, for EOCs such as temperature ($24^{\circ}\text{C} \leq T \leq 60^{\circ}\text{C}$), flow rate ($45.5 \text{ m}^3/\text{h}$ and $102 \text{ m}^3/\text{h}$), and inner pressure variation, operational noises, as well as flowing water inside the pipe.

In the training stage of this framework, the sparse discriminant vector is obtained so that the projection of the signals on this vector is a good predictor of class labels (*e.g.*, 1.0 for intact pipes, and -1.0 for damaged pipes). Every arrival in the time trace of the signal is assigned a coefficient representing their significance in predicting the class labels. Therefore, the arrivals with less contribution in defining the state of the pipe are assigned close-to-zero coefficients.

It is important to evaluate the performance of the SD method when training and test data are different in terms of parameters that can affect time-location of the damage-sensitive arrivals. These parameters include EOCs, as well as damage characteristics, such as location, type, size, and number of damage(s). I analyzed and experimentally validated the detection performance of the SD method under different scenarios. Detection accuracies above 96% have been achieved for all the experiments reported in this chapter.

Although the robustness of the SD method to training parameters has been validated, its application is limited to the cases where access to the pipe, to impose training abnormality, is possible and practical. Moreover, it is useful to reduce the number of training parameters that may affect detection performance. Therefore, in Chapter 5, I developed an *unsupervised* method

to address this limitation, in addition to the aforementioned challenges inherent in the nature of guided-waves.

Damage changes guided-wave signals. However, due to complex nature of these signals, and the EOC effects, in practice, detecting such changes is not trivial. The proposed idea in Chapter 5 is to extract a sparse subset of the arrivals that represents the majority of the energy of the signals for wide range of EOCs. Therefore, in such a subset, the effects of EOCs are less dominant compared to those in complete signals. The training data in the proposed approach is not labeled, but is assumed to mostly include the signals from the intact pipe. Due to the effects of damage, this subset will not be a good representative of such arrivals in the damaged signals. Therefore, the error in estimating signals' energy (ΔE) with this subset will increase when damage occurs. Moreover, occurrence of damage will increase the sensitivity of the ΔE values to EOC variations. These energy estimation errors (ΔE s) were used as damage-sensitive features in the proposed approach (namely, Sparse Energy (SE) method).

To validate the discriminatory power of the damage-sensitive features, and to investigate their distribution in both populations, I applied two heuristic classification methods , namely, a clustering method based on K-means, and a method based on support vector machine (SVM) with RBF kernel. For online detection purposes, however, a detection method based on Kullback-Leibler Divergence (KL) was applied. I validated the SE method using the same set of field and laboratory data used in Chapter 4. According to these results, the proposed SE method is successful in online detection of damage under varying EOCs. However, it is shown that the wider the range of EOCs in the training dataset, the better the detection performance of the SE

method. This range, however, is not required to be comprehensive of all possible testing scenarios. For example, for a test dataset in which temperature varied between 24°C and 39°C, a training dataset with temperatures ranging between 24°C and 30°C resulted in separation accuracy of 99%, and detection delay of five observations. Moreover, it was observed that detection performance could diminish as the training dataset includes damaged signals.

Figure 6.1 depicts the proposed framework for online damage detection of pipelines using the two approaches proposed in Chapter 4 and Chapter 5.

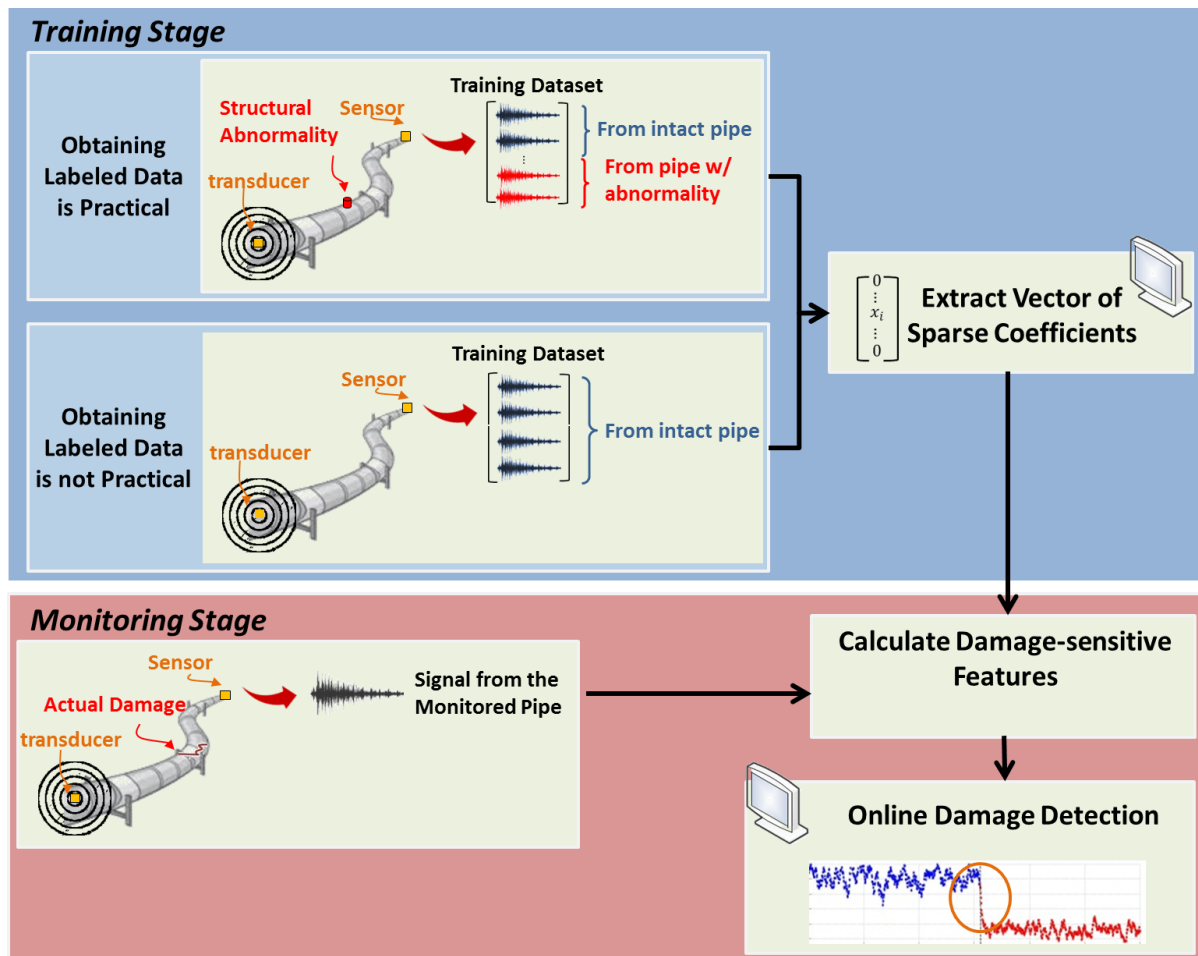


Figure 6.1: Overview of the proposed online damage detection framework.

6.2 Contributions

In Chapter 4 of this dissertation, I tested the following hypothesis:

A sparse subset of the arrivals in a diffuse-field guided-wave signal contains enough information for optimal damage detection, and this subset is less sensitive to EOC variations compared to the complete signal.

Extracting damage information from complex guided-wave signals has been the topic of vast amount of studies for over half a century. A number of these studies are based on compensating for particular EOCs and/or extraction of scatter signals. However, as discussed in Chapter 1, these methods are often ad hoc to a particular damage scenario, EOC, or wave mode, and therefore, lack extensibility to different realistic operational conditions.

To the best of my knowledge, the work presented in Chapter 4 is the first study that takes advantage of the sparsity of damage information in time domain of diffuse-field guided-wave signals for optimal detection of damages with different characteristics.

I have proved that the proposed method does not rely on comprehensive training dataset that covers the majority of EOC scenarios and/or requires prior knowledge about the damage characteristics (*e.g.*, type, location, size, number). Moreover, this study incorporates a formal understanding regarding the effects of several parameters on different aspects of the approach. I

believe that such physical and experimental intuitions are important for extending the approach to a variety of operational scenarios.

The robustness of the method to wide ranges of EOCs makes it attractive for online monitoring. The proposed framework has proved successful for online detection of small damages of different type, as well as occurrence of consecutive damages, under varying EOCs. The non-adoptive approach proposed here prevents the performance of the SD method to be affected by false alarms.

In addition to the advantages of the proposed detection approach, the extracted sparse subset of the signal provides important insight regarding the role of different arrivals in the signals' time trace. For example, I found that later arrivals and end-reflections could substantially contribute in estimating the state of the pipe. Moreover, my findings suggest that even small damages can affect the physical properties of the pipe, so that the first arrivals that do not include reflections from damage can contain significant damage information. Therefore, denoising and/or diagnosis methods that are based on removal of multi-path reflections and/or first arrivals may be vulnerable to eliminating useful information.

The most widely used solution to address the complexity of guided-waves is single-mode excitation with an array of transducers. The limitations of such approaches are discussed in Chapter 1. The proposed method in Chapter 4, however, uses multi-modal signals obtained from a simple low-cost, low-power, pitch-catch method. Therefore, besides the simplicity of the implementation, this method makes it possible to benefit from distinct characteristics of multiple

modes, since different wave modes are sensitive to different features of damage/EOCs. This further expands the application of the method.

In Chapter 5, an unsupervised method is proposed to address the potential challenges of the SD method in obtaining labeled data. Damage, in the form of variation in the pipe's wall thickness or surface roughness, alters guided-wave signals. However, differentiating such changes from the ones caused by EOC variations is not trivial, especially, in complex diffuse-field signals. In Chapter 5, I suggest that, for optimal damage detection, one do not need to study the complete set of signals' arrivals, but only a sparse subset of the arrivals that represents the majority of signals' energy, for wide range of EOCs.

For the method proposed in this chapter, the following facts are verified: (1) among the recorded samples, the effects of EOCs will be more significant on low-energy arrivals with small signal-to-noise ratios (SNR) compared to high-energy arrivals, and (2) damage changes the energy-content and/or time-location of the arrivals in guided-wave signals.

Based on the aforementioned facts, in Chapter 5, I tested the following hypotheses:

A sparse subset of the arrivals in a diffuse-field guided-wave signal represents the majority of the signal's energy, and this subset is less sensitive to EOC variations than the complete signal.

Due to the effects of damage, the sparse subset of the arrivals representing the majority of intact signals' energy is not a good representative of the energy of the damaged signals. Hence, such a sparse representation of intact signals is sensitive to occurrence of damage.

The experimental analysis reported in Chapter 5 verified substantial differences between the time locations of high-energy arrivals in intact signals and the ones in damaged signals. I have shown that such differences can be detected under a wide range of EOCs if only a sparse subset of the arrivals is considered. To the best of my knowledge, these facts have not been formally incorporated into damage detection methods. Moreover, this work is the first study taking advantage of the sparsity of the arrivals that represent the majority of the energy of a complex diffuse-field signal for damage detection under varying EOCs.

Similar to the SD method, detection capability of the SE method is not based on any assumption regarding a particular type, size and/or location of damage. That is, regardless of the nature of the physical phenomenon caused by damage (*e.g.*, mode conversion, change in phase velocity, *etc.*), if the extracted sparse subset is affected by damage, it can be detected. The proposed SE method has proved successful for detection of damages with different characteristics (*i.e.*, types, location, and size), under wide range of EOCs.

The aforementioned advantages of the SD method in implementation simplicity, benefitting from multi-modal waves, and high online detection performance apply to the SE method as well.

Moreover, my experiments suggest that: (a) The first set of high-energy arrivals has less contribution in total energy of the signals compared to the later arrivals. (b) Almost no overlap was observed between the SD and SE coefficients, suggesting that the arrivals with maximum damage information are the ones with least energy content. (c) Even small damages can affect the physical properties of the pipe. Such changes can affect the propagation of high-energy arrivals through undamaged sections of the pipe. Therefore, even the arrivals from undamaged sections may include damage information.

The two proposed frameworks provide a novel approach (Figure 6.1) supporting continuous online monitoring of pipelines operating under wide ranges of varying EOCs.

6.3 Limitations

The supervised online monitoring framework proposed in Chapter 4 has proved high detection performance for a wide range of EOCs and damage scenarios. The tradeoff for this level of performance, however, is the requirement for labeled training data. The unsupervised framework proposed in Chapter 5 eliminates the damage-related training parameters. However, this is achieved at the cost of slight increase in the sensitivity of the method to EOC variations. As discussed in Chapter 5, the detection performance can degrade if training data is obtained under very limited ranges of EOCs compared to the test data. It is, however, notable that this method outperforms the current state-of-the-art approaches that, for example, can only compensate for temperature variations of 0.5-1°C. The final framework proposed in this thesis (Figure 6.1)

provides a guideline for practitioners to choose the suitable approach for a particular application, by trading off the pros and cons of each approach.

6.4 Future Work

The online detection framework presented in this thesis is an initial stage for a complete online diagnosis approach. Below is the summary of the future research avenues for achieving this vision.

Damage Quantification and Characterization:

Development of damage in real life is a gradual phenomenon. It is important to detect and quantify the changes in the severity of damage throughout its gradual development. Moreover, work needs to be done to formally define the types of the defects that the proposed methods can detect. Intuitively, in both SD and SE methods, the observations from an in-progress damage can be distinguished as soon as the corresponding *sparse signals* deviate from those of undamaged pipe. As shown in the validation sections of Chapter 4 and Chapter 5, such deviations can be quite significant for a crack as small as 10% of the thickness of the pipe. In addition, these experiments suggest that the change in the severity of damage might be detectable. Future work includes expansion of these methods for detection of gradual development of defects.

Damage Localization:

For complete damage diagnosis, once occurrence of damage is detected, its precise location on the pipe should be identified. Signals scattered from damage, so called scatter signals, are usually

used for localization. However, extracting scatter signals from multi-modal guided-wave signals is not trivial. Our study (Eyμποosh et al., 2015) shows that this becomes even more challenging when multi-modal signals are exposed to varying EOCs. In Eyμποosh et al., (2015), I have simulated guided-waves propagating in a pipe under different temperatures. The results suggest that different wave modes have different sensitivities to temperature variations. Therefore, generally, under varying temperatures, scatter signal is nonlinearly related to other components of the signal. I applied a nonlinear feature extraction method (namely, nonlinear principal component analysis, NLPCA) to remove nonlinear dependencies between signal components (Eyμποosh et al., 2015). Therefore, in order to expand the work presented in this thesis to localization of the detected damages, one path is to use nonlinear methods for extracting scatter signals. Based on the findings reported in Eyμποosh et al. (2015), I believe that such methods will improve localization with the diffuse-field guided-wave signals.

Damage Detection of Buried Pipes:

Coupling material outside the pipe (*e.g.*, soil, water, *etc.*) can affect the propagation of the guided-waves, through energy leakage, mode conversion and/or additional mode generation. Evaluation of the proposed approaches for buried pipes was not within the scope of this thesis. However, since significant portion of the pipelines worldwide is buried, it is important to expand the proposed framework to such cases.

Bibliography

- Alleyne, D.N., Cawley, P., 1992. The interaction of Lamb waves with defects. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 39, 381–397.
- Alleyne, D.N., Lowe, M.J.S., Cawley, P., 1998. The Reflection of Guided Waves From Circumferential Notches in Pipes. *J. Appl. Mech.* 65, 635–641.
- Alleyne, D.N., Pavlakovic, B., Lowe, M.J.S., Cawley, P., 2001. Rapid, long range inspection of chemical plant pipework using guided waves. *AIP Conf. Proc.* 557, 180.
- Aristégui, C., Lowe, M.J.S., Cawley, P., 2001. Guided waves in fluid-filled pipes surrounded by different fluids. *Ultrasonics* 39, 367–375.
- Belanger, P., Cawley, P., 2009. Feasibility of low frequency straight-ray guided wave tomography. *NDT E Int.* 42, 113–119.
- Belouchrani, A., Amin, M.G., 1998. Blind source separation based on time-frequency signal representations. *IEEE Trans. Signal Process.* 46, 2888–2897.
- Cawley, P., LOWE, M., Alleyne, D.N., Pavlakovic, B., Wilcox, P.D., 2003. Practical long range guided wave inspection-applications to pipes and rail.
- Cawley, P., Lowe, M.J.S., Simonetti, F., Chevalier, C., Roosenbrand, A.G., 2002. The variation of the reflection coefficient of extensional guided waves in pipes from defects as a

- function of defect depth, axial extent, circumferential extent and frequency. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* 216, 1131–1143.
- Chang Liu, J.B.H., 2013. Ultrasonic Monitoring of a Pressurized Pipe in Operation 1903–1913.
- Clarke, T., Cawley, P., Wilcox, P.D., Croxford, A.J., 2009. Evaluation of the damage detection capability of a sparse-array guided-wave SHM system applied to a complex structure under varying thermal conditions. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 56, 2666–2678.
- Cohen, L., 1989. Time-frequency distributions-a review. *Proc. IEEE* 77, 941–981.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297.
- Croxford, A.J., Moll, J., Wilcox, P.D., Michaels, J.E., 2010. Efficient temperature compensation strategies for guided wave structural health monitoring. *Ultrasonics* 50, 517–528.
- Croxford, A.J., Wilcox, P.D., Konstantinidis, G., Drinkwater, B.W., 2007. Strategies for overcoming the effect of temperature on guided wave structural health monitoring, in: *SPIE*. p. 65321T–1.
- Davies, J., Cawley, P., Lowe, M.J.S., 2008. Long Range Guided Wave Pipe Inspection – the Advantages of Focusing. Presented at the 17th World Conference on Nondestructive Testing, Shanghai, China.
- Degtyar, A.D., Huang, W., Rokhlin, S.I., 1996. Wave Propagation in Stressed Composites, in: Thompson, D.O., Chimenti, D.E. (Eds.), *Review of Progress in Quantitative Nondestructive Evaluation*. Springer US, pp. 1669–1676.
- Dehghan-Niri, E., Salamone, S., 2014. A multi-helical ultrasonic imaging approach for the structural health monitoring of cylindrical structures. *Struct. Health Monit.*

- Demma, A., Cawley, P., Lowe, M., Roosenbrand, A.G., 2003. The reflection of the fundamental torsional mode from cracks and notches in pipes. *J. Acoust. Soc. Am.* 114, 611–625.
- Demma, A., Cawley, P., Lowe, M., Roosenbrand, A.G., Pavlakovic, B., 2004. The reflection of guided waves from notches in pipes: a guide for interpreting corrosion measurements. *NDT E Int.* 37, 167–180.
- Donoho, D.L., Huo, X., 2001. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inf. Theory* 47, 2845–2862.
- DOT, 2010. The State Of The National Pipeline Infrastructure. US Department Of Transportation.
- EPA, 2007. Distribution System Inventory, Integrity and Water Quality| US EPA.
- Eybpoosh, M., Bergés, M., Noh, H.Y., 2015. Nonlinear feature extraction methods for removing temperature effects in multi-mode guided-waves in pipes, in: *SPIE Smart Structures/NDE*. San Diego, California, United States.
- Eybpoosh, M., Bergés, M., Noh, H.Y., 2014a. Investigation on the effects of environmental and operational conditions (EOC) on diffuse-field ultrasonic guided-waves in pipes, in: *International Conference on Computing in Civil and Building Engineering*. Presented at the ICCCB14, Orlando, FL.
- Eybpoosh, M., Bergés, M., Noh, H.Y., 2014b. Toward characterizing the effects of environmental and operational conditions on diffuse-field ultrasonic guided-waves in pipes, in: *SPIE Smart Structures/NDE*. San Diego, California, United States.

- Grant, M.C., Boyd, S.P., 2008. Graph Implementations for Nonsmooth Convex Programs, in: Blondel, V.D., Boyd, S.P., Kimura, H. (Eds.), Recent Advances in Learning and Control, Lecture Notes in Control and Information Sciences. Springer London, pp. 95–110.
- Guerrero-Mosquera, C., Trigueros, A.M., Franco, J.I., Navia-Vázquez, Á., 2010. New feature extraction approach for epileptic EEG signal detection using time-frequency distributions. *Med. Biol. Eng. Comput.* 48, 321–330.
- Harley, J.B., Moura, J.M.F., 2014. Data-driven matched field processing for Lamb wave structural health monitoring. *J. Acoust. Soc. Am.* 135, 1231–1244.
- Harley, J.B., Moura, J.M.F., 2013. Sparse recovery of the multimodal and dispersive characteristics of Lamb waves. *J. Acoust. Soc. Am.* 133, 2732–2745.
- Harley, J.B., Moura, J.M.F., 2013. Broadband localization in a dispersive medium through sparse wavenumber analysis, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4071–4075.
- Harley, J.B., Ying, Y., Moura, J.M.F., Oppenheim, I.J., Sobelman, L., Garrett, J.H., 2012. Application of Mellin transform features for robust ultrasonic guided wave structural health monitoring. *AIP Conf. Proc.* 1430, 1551–1558.
- Huang, K., Aviyente, S., 2006. Sparse representation for signal classification, in: In Adv. NIPS.
- Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B.R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S.A., Schnell, R., Tans, P., Trainer, M., Zamora, R., Conley, S., 2013. Methane emissions

- estimate from airborne measurements over a western United States natural gas field. *Geophys. Res. Lett.* 40, 4393–4397.
- Konstantinidis, G., Wilcox, P.D., Drinkwater, B.W., 2007. An Investigation Into the Temperature Stability of a Guided Wave Structural Health Monitoring System Using Permanently Attached Sensors. *IEEE Sens. J.* 7, 905–912.
- Kullback, S., Leibler, R.A., 1951. On Information and Sufficiency. *Ann. Math. Stat.* 22, 79–86.
- Lefebvre, F., Deblock, Y., Campistron, P., Ahite, D., Fabre, J.J., 2002. Development of a new ultrasonic technique for bone and biomaterials in vitro characterization. *J. Biomed. Mater. Res.* 63, 441–446.
- Liu, C., Harley, J.B., Ying, Y., Oppenheim, I.J., Bergés, M., Greve, D.W., Garrett, J.H., 2013. Singular value decomposition for novelty detection in ultrasonic pipe monitoring. p. 86921R–86921R–11.
- Liu, C., Harley, J., O'Donoghue, N., Ying, Y., Altschul, M.H., Berges, M., Garrett, J.H., Greve, D.W., Moura, J.M.F., Oppenheim, I.J., Soibelman, L., 2012a. Robust change detection in highly dynamic guided wave signals with singular value decomposition, in: *Ultrasonics Symposium (IUS), 2012 IEEE International*. Presented at the Ultrasonics Symposium (IUS), 2012 IEEE International, pp. 483–486.
- Liu, C., Harley, J., O'Donoghue, N., Ying, Y., Altschul, M.H., Garrett, J., James H., Moura, J.M.F., Oppenheim, I.J., Soibelman, L., 2012b. Ultrasonic monitoring of a pipe under operating conditions, in: *International Society for Optics and Photonics*.
- Long, R., Cawley, P., Lowe, M., 2003a. Acoustic wave propagation in buried iron water pipes. *Proc. R. Soc. Lond. Ser. Math. Phys. Eng. Sci.* 459, 2749–2770.

- Long, R., Lowe, M., Cawley, P., 2003b. Attenuation characteristics of the fundamental modes that propagate in buried iron water pipes. *Ultrasonics* 41, 509–519.
- Løvstad, A., Cawley, P., 2012. The reflection of the fundamental torsional mode from pit clusters in pipes. *NDT E Int.* 46, 83–93.
- Lowe, M.J.S., Alleyne, D.N., Cawley, P., 1998a. The Mode Conversion of a Guided Wave by a Part-Circumferential Notch in a Pipe. *J. Appl. Mech.* 65, 649–656.
- Lowe, M.J.S., Alleyne, D.N., Cawley, P., 1998b. Defect detection in pipes using guided waves. *Ultrasonics, Ultrasonics International* 1997 36, 147–154.
- Lu, Y., Michaels, J.E., 2009. Feature Extraction and Sensor Fusion for Ultrasonic Structural Health Monitoring Under Changing Environmental Conditions. *IEEE Sens. J.* 9, 1462–1471.
- Lu, Y., Michaels, J.E., 2005. A methodology for structural health monitoring with diffuse ultrasonic waves in the presence of temperature variations. *Ultrasonics* 43, 717–731.
- Nagy, P.B., Simonetti, F., Instanes, G., 2014. Corrosion and erosion monitoring in plates and pipes using constant group velocity Lamb wave inspection. *Ultrasonics* 54, 1832–1841.
- Niethammer, M., Jacobs, L.J., Qu, J., Jarzynski, J., 2000. Time-frequency representation of Lamb waves using the reassigned spectrogram. *J. Acoust. Soc. Am.* 107, L19–L24.
- Nurmalia, Nakamura, N., Ogi, H., Hirao, M., 2013. Mode Conversion and Total Reflection of Torsional Waves for Pipe Inspection. *Jpn. J. Appl. Phys.* 52, 07HC14.
- PHMSA, 2012. Leak Detection Study (No. DTPH56-11-D-000001). U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration.

- PHMSA, n.d. Pipeline Incident 20 Year Trend. US DOT Pipeline and Hazardous Materials Safety Administration.
- Pierce, A.D., Kil, H.-G., 1990. Elastic Wave Propagation from Point Excitations on Thin-Walled Cylindrical Shells. *J. Vib. Acoust.* 112, 399–406.
- Protopappas, V.C., Fotiadis, D.I., Malizos, K.N., 2006. Guided ultrasound wave propagation in intact and healing long bones. *Ultrasound Med. Biol.* 32, 693–708.
- Raghavan, A., Cesnik, C.E.S., 2007. Studies on effects of elevated temperature for guided-wave structural health monitoring. p. 65290A–65290A–12.
- Rose, J.L., 2004. *Ultrasonic Waves in Solid Media*. Cambridge University Press.
- Scalea, F.L. di, Salamone, S., 2008. Temperature effects in ultrasonic Lamb wave structural health monitoring systems. *J. Acoust. Soc. Am.* 124, 161–174.
- Schulz, M.J., Sundaresan, M.J., Mcmichael, J., Clayton, D., Sadler, R., Nagel, B., 2003. Piezoelectric Materials at Elevated Temperature. *J. Intell. Mater. Syst. Struct.* 14, 693–705.
- Seco, F., Jiménez, A., 2012. Modelling the generation and propagation of ultrasonic signals in cylindrical waveguides, in: *Ultrasonic Waves*. pp. 1–28.
- Shinozuka, M., Chou, P.H., Kim, S., Kim, H., Karmakar, D., Lu, F., 2010. Non-invasive acceleration-based methodology for damage detection and assessment of water distribution system. *Smart Struct. Syst.* 6, 545–559.
- Shull, P.J., 2002. *Nondestructive Evaluation: Theory, Techniques, and Applications*. CRC Press.

- Siqueira, M.H.S., Gatts, C.E.N., da Silva, R.R., Rebello, J.M.A., 2004. The use of ultrasonic guided waves and wavelets analysis in pipe inspection. *Ultrasonics* 41, 785–797.
- Sun, Z., Mao, Y., Jiang, W., Zhang, D., 2000. Investigation on interaction of Lamb waves and circumferential notch in pipe by means of wavelet transform, in: 2000 IEEE Ultrasonics Symposium. Presented at the 2000 IEEE Ultrasonics Symposium, pp. 827–830.
- Tse, P.W., Wang, X., 2013. Characterization of pipeline defect in guided-waves based inspection through matching pursuit with the optimized dictionary. *NDT E Int.* 54, 171–182.
- Vanlanduit, S., Parloo, E., Cauberghe, B., Guillaume, P., Verboven, P., 2005. A robust singular value decomposition for damage detection under changing operating conditions and structural uncertainties. *J. Sound Vib.* 284, 1033–1050.
- Wang, X., Tse, P.W., Mechefske, C.K., Hua, M., 2010. Experimental investigation of reflection in guided wave-based inspection for the characterization of pipeline defects. *NDT E Int.* 43, 365–374.
- Worlton, D.C., 1957. *Non-Destructive Testing* 15.
- Yeum, C.M., Sohn, H., Lim, H.J., Ihn, J.B., 2014. Reference-free delamination detection using Lamb waves. *Struct. Control Health Monit.* 21, 675–684.
- Ying, Y., Garrett, J., Harley, J., Oppenheim, I., Shi, J., Soibelman, L., 2013. Damage Detection in Pipes under Changing Environmental Conditions Using Embedded Piezoelectric Transducers and Pattern Recognition Techniques. *J. Pipeline Syst. Eng. Pract.* 4, 17–23.