Data-driven semi-supervised and supervised learning algorithms for health monitoring of pipes

Debarshi Sen[a], Amirali Aghazadeh[b], Ali Mousavi[b], Satish Nagarajaiah[a,c,*], Richard Baraniuk[b], Anand Dabak[d]

[a] *Department of Civil and Environmental Engineering, Rice University, Houston, TX 77005, USA*
[b] *Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005, USA*
[c] *Department of Mechanical Engineering, Rice University, Houston, TX 77005, USA*
[d] *Texas Instruments, Dallas, TX, USA*

Abstract

The use of guided ultrasonic waves (GUWs) for SHM of pipelines has been a popular method for over three decades. The superiority of GUWs over traditional vibration-based techniques lie in its ability to detect small damages (cracks and corrosion) over a satisfactory length of a pipeline. The physics of the system, however, is extremely involved that renders model-based techniques computationally prohibitive. Data-driven approaches, based on statistical learning algorithms, are far more suitable in such scenarios. In this paper, we propose two data-driven techniques, involving a semi-supervised and a supervised learning approach, for damage detection in pipes. In addition to circumventing the use of a model-based approach, the proposed approaches also aid in reducing the number of sensors deployed, leading to reductions in maintenance costs. The semi-supervised learning-based approach detects the presence of damage using a hierarchical clustering-based algorithm. The supervised learning-based approach performs damage localization in a multinomial logistic regression framework. We validate the proposed algorithms by acquiring guided wave responses from experimental pipes in a pitch-catch configuration using low-cost piezoelectric transducers. We demonstrate that our fully data-driven techniques accurately detect and localize cracks on two cast iron pipes of different lengths using a combination of two sensors.

Keywords: Data-driven structural health monitoring, Damage detection, Wave propagation in pipes, Hierarchical clustering, Multinomial logistic regression

1. Introduction

Structural Health Monitoring (SHM) is essential for maintenance of civil infrastructure and aerospace systems. It entails detection of damages in systems that helps provide estimates of remaining useful life [1]. Typically, damage detection is classified into four levels [2], namely levels I to IV. In order of increasing effort, required for accomplishment, these levels are: detection, i.e., whether or not a damage has occurred, localization, i.e., where the damage is, quantification, i.e., how severe the damage is, and service life estimation, i.e., how long will the system be operational.

SHM of pipes in particular, has been a crucial task for engineers for the past quarter of a century owing to the catastrophic nature of failures, not only for the environment but the society as a whole [3]. Guided ultrasonic wave (GUW)-based technology has been in use for SHM of pipelines for many years now [4, 5]. The readers can find a review of these applications in Liu and Kleiner [6] which classifies the application of GUW-based SHM as either passive or active sensing. As the names suggest, active sensing involves

*Corresponding author
Email addresses: debarshi.sen@rice.edu (Debarshi Sen), amirali@rice.edu (Amirali Aghazadeh), ali.mousavi@rice.edu (Ali Mousavi), satish.nagarajaiah@rice.edu (Satish Nagarajaiah), richb@rice.edu (Richard Baraniuk), dabak@ti.com (Anand Dabak)

actuation of the systems at hand using an external source, whereas, passive sensing involves acquisition of signals resulting from ambient vibrations. In this paper, we focus on active sensing.

A survey of the literature on pipeline monitoring using GUW-based active sensing yields ample research. These range from development of sensors to algorithms for crack and corrosion detection. For example, Rose [7, 8] uses GUWs for crack detection in pipelines. Giurgiutiu [9] uses piezoelectric wafer active sensors [10] for corrosion detection on pipelines. Lee et al. [11] propose a baseline-free pipeline monitoring scheme taking aid of optical fiber-guided laser ultrasonics. Na and Kundu [12] study the effects of varying incident angles of ultrasonic waves for defect detection in underwater pipelines. Lowe et al. [5] provide a comprehensive discussion on this approach to damage detection in pipelines. In addition to the efforts of the research community, major industrial sectors plan on enhancing their capabilities as well with the market size expected to grow from USD 4.13 Billion in 2015 to USD 8.72 Billion by 2026 [13].

Traditional approaches to damage detection in pipelines typically involve the use of model-based approaches wherein a high-fidelity model represents system behavior [14, 15] and hence allows for detecting changes in the system due to advent of damage. However, constructing such models is in many instances physically intractable and computationally prohibitive. Given the advancement in sensors and hardware development, data-driven approaches become far more suitable for damage detection. The aim of data-driven techniques is to alleviate the need of a high-fidelity model for differentiating between damaged and undamaged configurations of the system. The past decade has seen a rise in attention towards data-driven techniques for SHM using GUWs [16, 17, 18, 19]. Some of the popular efforts in pipeline monitoring are from Liu et al. [20], Ying et al. [21, 22] and Eybpoosh et al. [23]. Data-driven approaches based on statistical learning algorithms typically require extensive signal processing of raw data for feature selection. Such an approach helps improve the efficiency of the damage detection algorithms in the literature. In addition, typically the existing methods require a moderately dense to highly dense array of sensors for efficient performance. In this paper we address two key issues; first, we circumvent the feature selection phase and demonstrate that bandpass filtered signals themselves are efficient for damage detection, second, we minimize the number of sensors used for performing GUW-based damage detection. The second focus of the paper also helps in minimizing maintenance costs as we show that the proposed algorithms perform efficient damage detection using low-cost ($5 per sensor) piezoelectric transducers.

To accomplish these objectives, we propose two approaches involving a semi-supervised and a supervised statistical learning algorithm. The semi-supervised approach performs level I damage detection, whereas the supervised one performs damage localization or level II damage detection. We test the efficacy of the proposed algorithms experimentally on cast iron pipes with bituminous anti-corrosion paint layer. We demonstrate that effective damage detection is possible with just two low-cost piezoelectric sensors using the proposed semi-supervised learning scheme. In addition, we demonstrate reliable damage localization using the supervised learning framework.

We organize the paper as follows: First, we describe the various statistical learning algorithms that we utilize. Next, we describe the elastic wave propagation physics that we harness for making the proposed algorithm work, following that we describe the proposed approaches. Finally, we describe the experimental setup and discuss the performance of the proposed approaches.

2. Statistical learning algorithms background

In this section, we present the statistical algorithms we use in our proposed approach. We briefly describe principal component analysis (PCA), hierarchical clustering (HC) and multinomial logistic regression (MLR). For a thorough discussion on these algorithms, we direct the readers to the book by Hastie *et al.* [24]. Statistical learning algorithms are typically of two types: supervised and unsupervised. Supervised algorithms require data sets with corresponding data labels for each sample. Meta models are then generated using these data sets. Subsequently, these models are used on data acquired later for label prediction. Unsupervised algorithms on the other hand, do not use data labels but simply look at datasets directly. Although label prediction is not possible using such algorithms, one may use these algorithms for observing patterns in data. PCA and HC are unsupervised algorithms, whereas, MLR is a supervised scheme.

2.1. Principal component analysis

Principal component analysis (PCA) is an unsupervised statistical learning technique. It typically aids in dimensionality reduction and visualization. If $\mathbf{X} \in \mathbb{R}^{n \times p}$ is a data set (where each column is centered, i.e., each column has zero mean) such that, n is the number of observations and p is the number of features, the covariance matrix of the data is defined as $\mathbf{X}^T\mathbf{X}$. The eigenvectors of the covariance matrix are the principal component directions, defined as,

$$\mathbf{X}^T\mathbf{X} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T, \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{p \times p}$ is the matrix whose columns are the eigenvectors of the covariance matrix and $\mathbf{\Lambda} \in \mathbb{R}^{p \times p}$ is a diagonal matrix, whose diagonal elements are the corresponding eigenvalues. The projected version of the data in principal component space is

$$\hat{\mathbf{X}} = \mathbf{X}\mathbf{\Phi}. \tag{2}$$

The above computation can also be performed using a Singular Value Decomposition (SVD) approach. The SVD of the data matrix is defined as:

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{3}$$

where $\mathbf{U} \in \mathbb{R}^{n \times n}$ and $\mathbf{V} \in \mathbb{R}^{p \times p}$ are unitary matrices and $\mathbf{D} \in \mathbb{R}^{n \times p}$ is a diagonal matrix whose diagonal elements are the corresponding singular values sorted in descending order. On applying this decomposition, the covariance matrix will be:

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{D}^T\mathbf{U}^T\mathbf{U}\mathbf{D}\mathbf{V}^T = \mathbf{V}\mathbf{D}^T\mathbf{D}\mathbf{V}^T. \tag{4}$$

By comparing equations 1 and 4, one can conclude that the matrix consisting the right singular vectors as columns, $\mathbf{V}$ is equivalent to the matrix of eigenvectors $\mathbf{\Phi}$ of the covariance matrix, $\mathbf{X}^T\mathbf{X}$ and the singular values of $\mathbf{X}$ are square roots of the eigenvalues of $\mathbf{X}^T\mathbf{X}$. This implies that the right singular vectors are the principal components and hence, the projected data on the principal component space is

$$\hat{\mathbf{X}} = \mathbf{X}\mathbf{V} = \mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V} = \mathbf{U}\mathbf{D}. \tag{5}$$

Each column of the matrix $\hat{\mathbf{X}}$ represents the projection of the data set $\mathbf{X}$ on to each of the principal components. The projection of the data on the ith principal component can then be evaluated as $\mathbf{u}_i\sigma_i$, where $\mathbf{u}_i$ is the ith left singular vector (i^{th} column of $\mathbf{U}$) and σ_i is the ith singular value.

In this paper, we use PCA for:

- Dimensionality reduction. This allows for reducing the number of variables involved with signal and aids in focusing on the low rank information of the signals that are of interest to us.
- Data visualization. The data sets we acquire in this work are typically high-dimensional. For ease of visualization, we use two principal components as basis vector of a two dimensional plane, where we project our data for ease of visualization.

2.2. Hierarchical clustering

Clustering techniques are another unsupervised statistical learning approach. As the name suggests these algorithms congregate *similar* data into clusters. Various distance measures define the similarity of data. The number of clusters is an input parameter for these algorithms. In this paper, since we are interested in level I damage detection we set this parameter to two representing either a damaged or an undamaged state. The most basic algorithm for clustering is the k-means clustering. However, it is inefficient with nonuniform datasets and when within cluster distributions are not spherical [24]. Hence, we utilize hierarchical clustering (HC) in this paper.

HC typically constructs a hierarchy of data clusters. HC uses a dissimilarity function for allocating data points to clusters. Euclidean distance and Mahalanobis distance, to name a few, are popular choices for a dissimilarity function. The complete linkage variant of HC uses a minimax formulation for constructing the clusters based on the choice of dissimilarity function (other variants being single and median linkage).

Let there be two clusters $\mathbf{s}_j$ and $\mathbf{s}'_j$. Then the dissimilarity function for a complete linkage and the ensuing optimization problem for HC is:

$$\begin{aligned} d_{\text{complete}}(\mathbf{s}_j, \mathbf{s}'_j) = & \max_{i \in \mathbf{s}_j, i' \in \mathbf{s}'_j} d(i, i') \\ & \min_{\mathbf{s}_j, \mathbf{s}'_j} \quad d_{\text{complete}}(\mathbf{s}_j, \mathbf{s}'_j) \end{aligned} \tag{6}$$

The above equation essentially means in order to construct clusters, HC with complete linkage minimizes the maximum dissimilarity between two data points in the feature space for all clusters. In this paper, we use the Euclidean distance as a dissimilarity function. Given the number of required clusters, the algorithm reports the clusters based on the number of branches from the dendrogram associated with HC.

2.3. Multinomial logistic regression

Multinomial logistic regression is a supervised statistical learning approach for data classification. Given a set of training data and their K corresponding classes (labels) $\{c_k | k = 1, 2, \ldots, K\}$ the MLR algorithm learns a linear classifier that classifies a new data point into one of the K classes. The MLR model, assumes that a linear predictor function $f(i, k)$ predicts the probability that the data point $\mathbf{x}_i$ belongs to the class c_k. The linear predictor is parametrized by regression parameters $\boldsymbol{\beta}_k$ following the equation $f(i, k) = \mathbf{x}_i^T \boldsymbol{\beta}_k$. With this notation the probability of $\mathbf{x}_i$ being in class c_k can be derived as

$$P(Y_i = k) = \frac{e^{\boldsymbol{\beta}_k \mathbf{x}_i}}{1 + \sum_{p=1}^{K} e^{\boldsymbol{\beta}_p \mathbf{x}_i}}. \tag{7}$$

The unknown regression parameters $\boldsymbol{\beta}_k$ are typically jointly estimated in a high dimensional problem by maximum a posteriori (MAP) estimation during the training phase using the following optimization problem,

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{D} -\log\, p(Y_i | \mathbf{x}_i; \boldsymbol{\beta}) + \gamma \|\boldsymbol{\beta}\|_1, \tag{8}$$

where $\|\boldsymbol{\beta}\|_1$ is a regularization term promoting sparsity in the features.

3. Helical guided ultrasonic waves

In this section, we discuss the physical basis of the proposed algorithm. Since guided waves are extremely useful for SHM [25], the intricacies involved in the mechanics of wave propagation in a pipe has been addressed in great detail in the literature. The key difference between propagating elastic waves in plates and pipes is the different types of *wave modes* propagating. For studying wave propagation, we typically treat pipes as a three dimensional structure (either in Cartesian or polar coordinate systems) in contrast to plates which we generally model as two dimensional systems. Consequently, for plates two types of wave modes exist, namely, symmetric and anti-symmetric. Correspondingly, for pipes, we have three kinds of wave modes, namely, longitudinal, flexural and torsional. Presence of three kinds of waves, instead of two like in plates, makes wave propagation in pipes far more complicated. This can be observed when comparing the dispersion curves for plates and pipes.

For hollow pipes, a high diameter to pipe thickness ratio, leads to the propagating Lamb waves in a pipe behaving like those in a plate [26]. This implies that the wave propagation in the pipe can then be represented in terms of symmetric and antisymmetric wave modes similar to a plate. In such cases, wave propagation along the pipe can be thought of as helical waves along the surface of pipes [27] (see figure 1). Between any two points on the pipe, there are infinitely many possible helical paths joining them. If there is a damage in the pipe, few of the helical paths will be affected. If these waveforms are studied, the changes can be monitored for the purposes of damage detection. In the presence of a small damage, these changes are generally not profound enough for visual identification from response signals. Figure 2 shows the differences

in response between damaged and undamaged scenarios. The damaged and undamaged pipe responses are acquired from a pitch catch configuration deployed on a cast iron pipe described later in section 5. We show the envelope of the mean signals for the damaged and the undamaged scenarios along with the associated statistical uncertainties in figure 2. To simulate a damaged scenario, we use a mass as described in section 5. In this paper, we use helical GUWs in conjunction with statistical learning algorithms and demonstrate its use for damage detection in pipes.

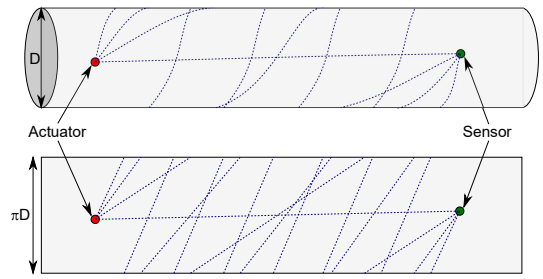


Figure 1: Helical guided ultrasonic waves as defined in [27]. A few of the multiple possible wave paths possible from the actuator and sensor are shown on the pipe the corresponding plate representation of the pipe. A damage typically blocks off a small number of such infinite possible paths for waves to reach the sensor. We use the small change in the acquired signal due to this change for the purpose of damage detection.

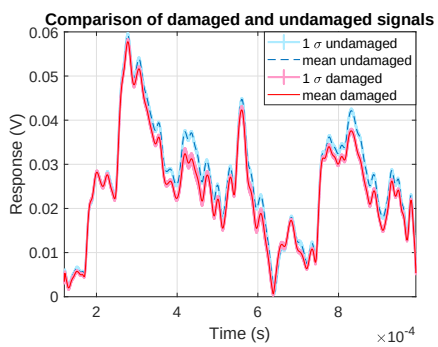


Figure 2: A comparison between envelope of mean damaged and mean undamaged response signals, along with a measure of the variance, acquired from a real pipe. Damage is simulated by attaching a mass to the pipe.

4. Proposed algorithms

As discussed earlier, we propose a semi-supervised and a supervised algorithm for damage detection and localization in pipes respectively. In this section, we discuss both the algorithms.

4.1. Semi-supervised approach

For the purposes of level I damage detection one would want to cluster the data into two clusters only, namely damaged and undamaged. As discussed earlier, HC is an unsupervised algorithm implying the absence of labels accompanying the data. Hence, without *a priori* information on the labels, classification of each cluster obtained from HC is not possible. Hence, we require a single data from the undamaged state of a pipe as input to this algorithm with a label. Once, a label is available for one of the undamaged signals, the cluster consisting of the labeled data may be classified as constituting the undamaged responses, and the other damaged. The proposed algorithm may be summarized as follows:

- Step 1: Acquire response signals (damaged or undamaged) from a pitch-catch set up.

- Step 2: Apply hierarchical clustering with complete linkage and Euclidean distance as the difference measure with number of clusters set to two.
- Step 3: Search for the labeled undamaged data in the clusters.
- Step 4: Assign the cluster with the undamaged label where the labeled data is present. This essentially leads to the knowledge about existence of damage.

4.2. Supervised approach

In order to identify the location of the damage on the pipe we formulate a logistic regression classification problem where the data points are the profiles of the signal as the result of a damage occurring on a region of the pipe from which we obtain the classes. We start by segmenting the pipe into K different regions along its length and train a logistic classifier to predict the region of damage.

Our main challenge in applying the classical logistic regression classifier is the dimensionality of the data points. In our experiments the dimension of signal from a damaged scenario is significantly bigger than the number of damaged data points we could obtain due to various sensing and experimental limitation. In order to reduce the dimensionality of the data we use two common techniques in statistical learning: 1) Random projection which reduces the dimension of the data by multiplying the data vectors by a matrix whose entries are drawn from random Gaussian distribution, and 2) Principle components analysis which project the data vectors into a lower dimensional space defined by the directions in the space that explain the variance of the data, the most. We use both techniques in order to reduce the dimensionality of the data and find that PCA outperforms random projections.

5. Experimental setup

5.1. Description of the system and input excitation

We use two cast iron pipes for our experiments. Table 1 lists a brief description of each pipe. Both the pipes have a bituminous anti-corrosive paint on the surface. This paint layer typically attenuates the applied excitation by affecting the contact between the actuator and the pipe. Similarly, the signals we acquire from the sensors are also further attenuated by this layer. However, the degree of attenuation is unknown. Additionally, it is also unknown whether the paint layer filters the signals in any other way. This is similar to a real life scenario where the properties, and their effect on acquired signals, of paints on existing pipeline networks may be unknown. However, as we will show in the results section, the proposed algorithm is effective even in presence of such a paint layer.

Pipe label	Material	Length (m)	Outer diameter (cm)	Thickness (mm)
A	Cast iron	1.23	16	5.2
B	Cast iron	3.05	16	6.73

Table 1: Description of each of the pipes used for the experiments. The labels for each pipe listed here will be used henceforth in this report to refer to each pipe

The pipes are supported at the ends on wooden benches with rubber padding to isolate it from external vibrations from the laboratory environment. For acquiring damaged response, we simulate presence of damage and do not actually damage the pipe. We attach a mass on the pipe at the desired damage locations. To ensure acoustic coupling between the mass and the pipe, we use grease. The mass scatters the elastic waves propagation in body of the pipe (similar to a real damage).

The placement of the sensors is as shown in figure 3. The location of the actuator and sensor divides the length of the pipe into three regions. We assume the actuator is always along the zenith line of the pipe, i.e., at an offset of zero degrees in view A-A in figure 3. We define the set of angles with respect to the zenith line at which the sensors may be attached as $\theta_S \in \{0°, 30°, 60°, ..., 180°\}$. Similarly, the angular position of the damage, i.e., the mass, may be either of those angular positions. The choice of these angles will shed light on the performance of the proposed approach where there is an angular offset between the relative angular

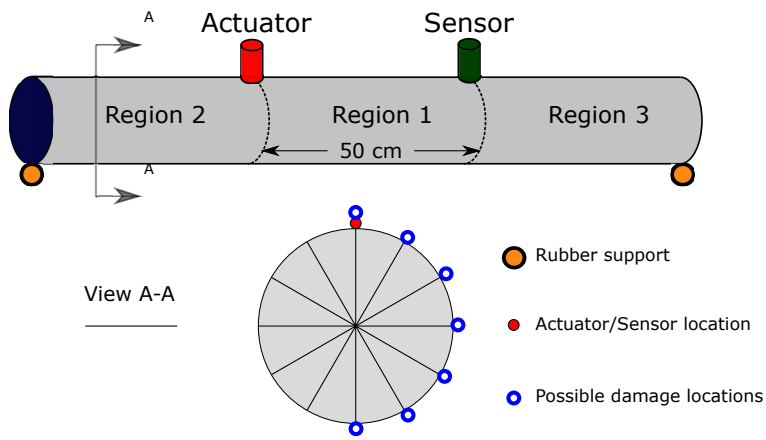


Figure 3: The general setup of the pipes. Based on the position of the actuator and sensor, we divide the pipe into three regions in the vicinity of the actuator-sensor pair. The sectional view A-A shows the possible angular positions of the sensors and damages with respect to the angular position of the actuator.

positions of the sensor and the damage with respect to the actuator. For pipe B we use the same setup in two different regions. One is away from the ends and the other near the ends. We perform the experiments near the ends too to observe the impact of boundary effects on the proposed approach.

As discussed earlier, we focus on guided wave based active damage detection. The external excitation we use is a five cycle tone burst signal such that the frequency spectrum of the signal is narrow. This is to ensure minimal wave dispersion. We select the central frequency of the tone burst signal such that the number of wave modes propagating in the elastic medium is minimized in order to reduce the complexity of the acquired signals. Figure 4 shows a typical input excitation we use, in the time and frequency domain.

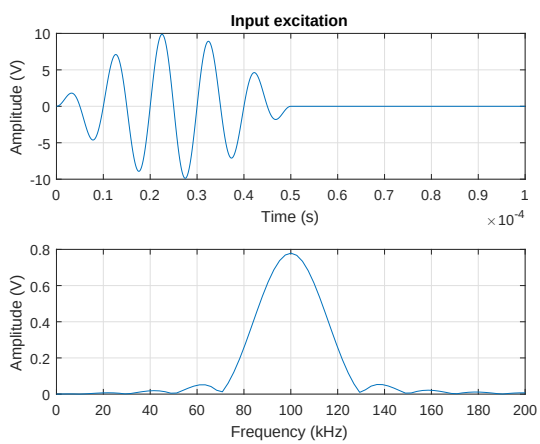


Figure 4: A typical 5-cycle tone burst input excitation with a central frequency of 100 kHz.

5.2. Hardware used

We use circular disk shaped lead zirconium titanate (PZT) piezoelectric sensors manufactured by Steminc piezo in all our experiments. These sensors are used as both the actuator (transmitter) as well as sensor (receiver). They are 10 mm in diameter and 3 mm thick with dominant R mode of vibration with a resonant frequency of 215 kHz. We attach them on pipes A and B using liquid nails, a glue manufactured by perfect glue. This ensures that the angle of incidence of the pressure waves are 90°, resulting in propagation of both symmetric and anti-symmetric wave modes along the pipe. We use a Tektronix AFG3022B and

Tektronix TDS2025C with four channels as the function generator and oscilloscope respectively. We use Texas Instruments (TI) high frequency booster packs (provided by TI) for amplification of received signals. We perform denoising by averaging one hundred and twenty eight copies of the acquired signals and bandpass filtering them. Figure 5 shows the experimental setup described above.

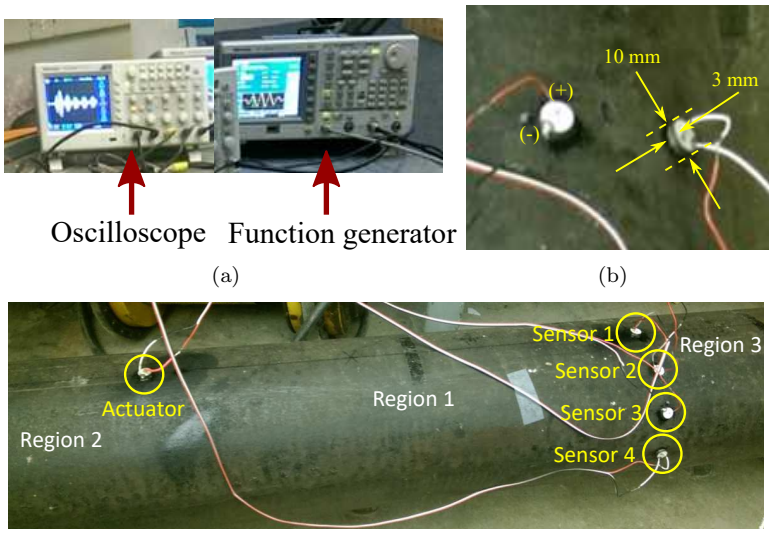


Figure 5: The experimental setup. (a) The oscilloscope and the function generator used for the experiments, (b) the sensors deployed on the pipes, (c) the three different regions shown with respect to the position of the actuator and sensor. We have one actuator and four sensors at four angular locations.

5.3. Data acquisition

For the setup shown in figure 3, we acquire data from all the three regions for both damaged and undamaged cases. We use an average of 128 acquired signals to improve the signal to noise (SNR) of the acquired signal. We perform further signal de-noising by applying a bandpass filter on the signals. We design the bandpass filter such that it has a spread of 100 kHz around the central frequency of the pulse used for actuation. As discussed earlier, we simulate damage by using a mass which scatters the high frequency waves that propagate through the structural medium of the pipe. For pipe A, we collect 20 samples per region and for pipe B we acquire 100 samples. Figure 6 shows a typical undamaged and damaged signals acquired from pipe B when the damage is in region 1, with the mass being placed at 0°.

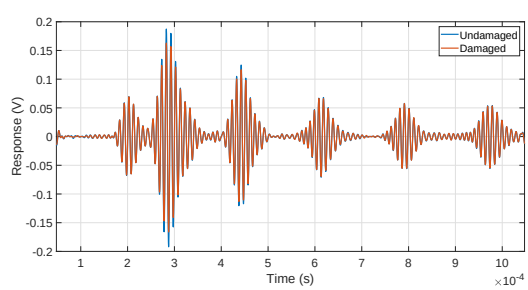


Figure 6: A typical undamaged and damaged signal acquired forom pipe B, when the damage is in region 1 and at 0°.

6. Results: Semi-supervised approach

6.1. Pipe A

We use results from pipe A to demonstrate the application of the semi-supervised portion of the proposed algorithm only. We avoid acquiring data from near the boundaries for pipe A. We will explore the effects of boundaries using the data from pipe B. As discussed earlier, there are six possible angular locations of damage, in each of the three different regions in figure 3.

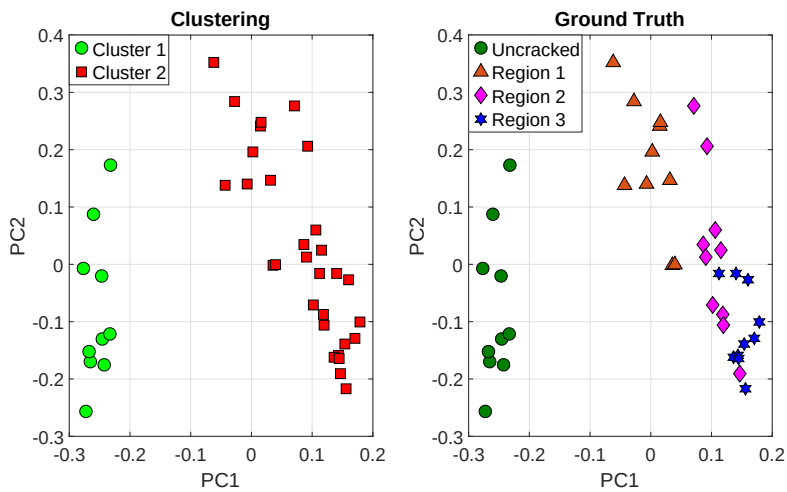


Figure 7: Results obtained from the proposed semi-supervised algorithm when the actuator, sensor and the simulated damage are all at an angular position of 0 degrees. The figure on the right shows the ground truth, i.e., the type of each data point, whether they are damaged or undamaged. The figure on the left shows the clusters obtained from HC. Clearly, for any of the undamaged data points treated as the *a priori* data, the proposed semi-supervised algorithm will provide a 100% accuracy in damage detection.

Figure 7 shows the results for the case where damage aligns with the straight line joining the actuator and the sensor. Both plots in the figure show a two dimensional representation of the data. The readers should note that each point shown is a time series in itself and is represented on a tow-dimensional plane. The first two principal component directions for the data form the basis for the two dimensional representation. The right hand side plot shows the ground truth, namely, it shows the source of each data point, i.e., undamaged or damaged from regions one, two and three. The clustering algorithm efficiently separates damaged and undamaged data into two separate clusters. Based on the proposed algorithm, we may treat any one of the undamaged points as the a priori data. Based on the clusters obtained, we can state that cluster 1 in figure 7 is a cluster of undamaged data points.

Figure 8 shows the dendrogram obtained from the hierarchical clustering algorithm. It highlights the differences in between various data points based on the dissimilarity function used for hierarchical clustering.

Figure 9 shows the clustering results for all the other angular positions of damage. Clearly, for angular damage positions between 0 and 90 degrees, the clustering algorithm is able to distinguish between the damaged and undamaged data points by comparing the clusters with the ground truth. We attribute the 100% efficiency to the strong scattering signals from the damage acquired by the sensors. For larger angular offsets, the signal strength of scattered waves reaching the sensor reduces, that leads to reduction of classification accuracy.

In figure 9 (d), for angular damage position 120°, we observe a few damaged data included in cluster 1, which consists of all the undamaged data. Data points with red circles around them in figure 9 (d) shows such points. This implies that some damaged data points are classified as undamaged data points. Similarly, figure 9 (e) shows the results for angular damage offset of 150°. The blue circles show undamaged points clustered into cluster 2 that consists of all of the damaged points. This leads to a situation wherein, if any

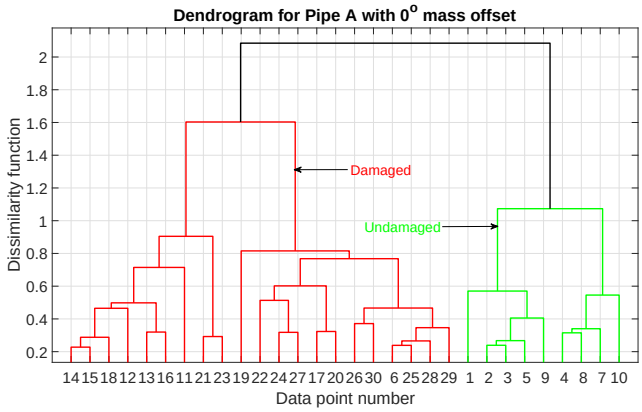


Figure 8: A dendrogram obtained from hierarchical clustering for the results shown in figure 7. One can observe how each data point has been clustered based on the Euclidean dissimilarity function. The labels "Damaged" and "Undamaged" are based on a priori undamaged data knowledge

one of the undamaged data from cluster 2 are selected as the a priori undamaged data, this will imply all the damaged data will be classified as undamaged.

To quantitatively define the performance of the proposed algorithm, we use statistical measures from hypothesis testing. We define the positive event as the detection of damage and no detection as negative. Hence, a true positive (TP) scenario is when a damage is detected when a damage exists in reality. Similarly we define, false positive (FP), true negative (TN) and false negative (FN) scenarios. Table 2 summarizes these definitions.

Term	**Ground Truth**	**Classified as**
True Positive (TP)	Undamaged	Undamaged
False Negative (FN)	Undamaged	Damaged
False Positive (FP)	Damaged	Undamaged
True Negative (TN)	Damaged	Damaged

Table 2: Definition of statistical measures used for quantifying the performance of the proposed semi-supervised learning algorithm.

To evaluate rates of each statistical measure, for each damage angular offset we evaluate the number of correct and incorrect classifications. This is done by considering each undamaged data as the a priori undamaged data. Figure 10 shows the performance of the proposed algorithm for various angular damage offsets. From the above results, we conclude that using a single actuator sensor pair we can accurately detect the presence of damage up to damage offsets of 120°. Since the accuracy is 100% up to an offset of 90°, we propose the deployment of single actuator-sensor pair units at diametrically opposite locations on the pipe surface to maintain 100% accuracy in damage detection. Figure 11 shows the proposed setup. The actuator-sensor pair on the top will cover the area shaded in red whereas the actuator-sensor pair at the bottom covers the area shaded in blue.

6.2. Pipe B

In this section we discuss the results from pipe B. Due to the relatively longer length of pipe B with respect to pipe A, we study the performance of the proposed semi-supervised approach at two different locations, one away from the boundary and one close to the boundary.

6.2.1. Away from the boundary

For pipe B, the setup of actuator and sensor pair is similar to that of pipe A. However, in this case, we acquire data from four sensors at once, to also analyze the effects of sensor locations. Figure 5 shows the

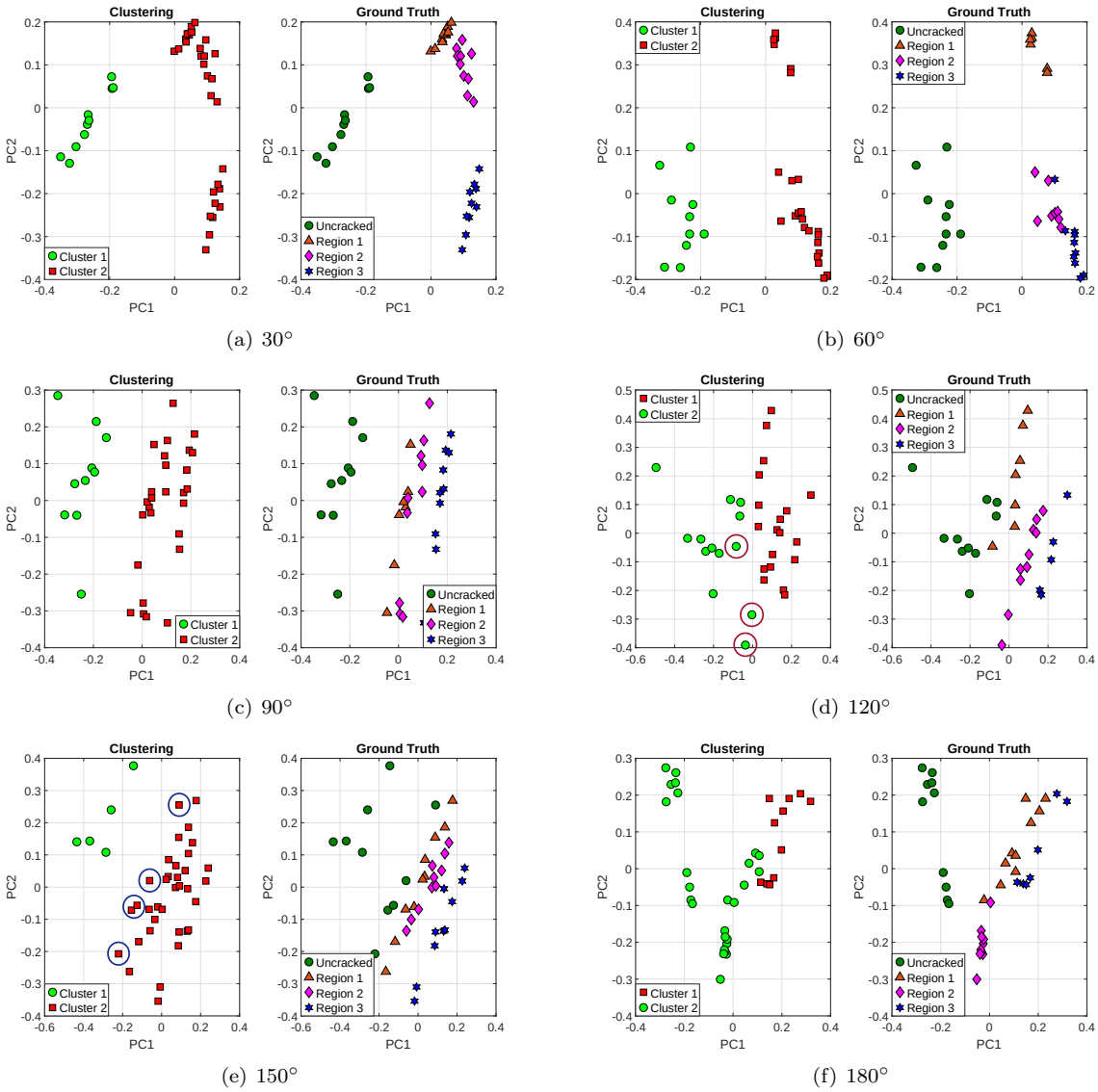


Figure 9: Results obtained from the proposed algorithm for various damage angular positions. In figure (d) the red circles show damaged data classified as undamaged and in figure (e) the blue circles show undamaged data classified as damaged.

locations of the actuator, the sensors and various regions, created with respect to actuator-sensor locations, of the pipe. The angular positions of the sensors are 0°, 30°, 60° and 90°.

We use the same input excitation as before. However, one should note that the thickness of pipe B is larger than that of pipe A. Hence, we expected a stronger signal attenuation leading to poorer performance with respect to pipe A. However, the thickness to diameter ratio is still within the limits of helical GUW assumption, hence the proposed semi-supervised approach is still applicable.

Figure 12 shows the results when the simulated damage is at an angular position of zero degrees with respect to the actuator. We demonstrate the results for two different angular positions of the sensor, zero and sixty degrees. Clearly for a damage location of zero degrees, the damaged and undamaged data are well separated. This again, is the best case scenario with minimum signal attenuation, hence produces the best results. The same damage location, but with a sensor that is offset by an angle of sixty degrees, produces worse results. Due to the relative distances in between the damage and the sensor location, it becomes more

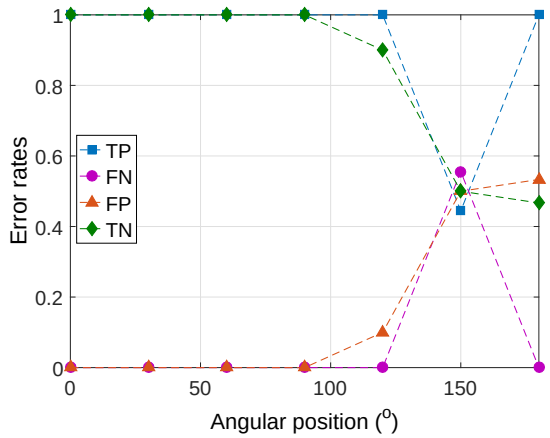


Figure 10: Statistical false alarm performance of the proposed semi-supervised algorithm for pipe A. It achieves 100% accuracy upto damage angular positions of 90°.

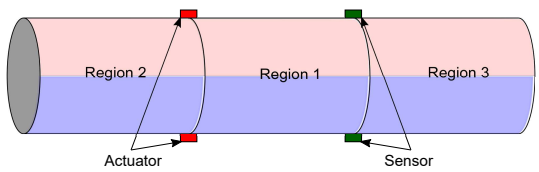
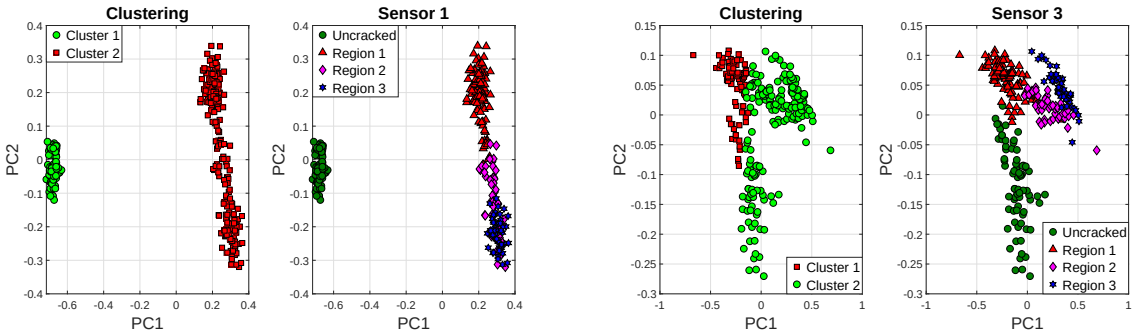


Figure 11: The proposed deployment of two actuator-sensor pairs for maintaining 100% accuracy in damage detection for all damage offsets. The he pair on the top covers the area shaded in red, while the pair at the bottom covers the area shaded in blue.



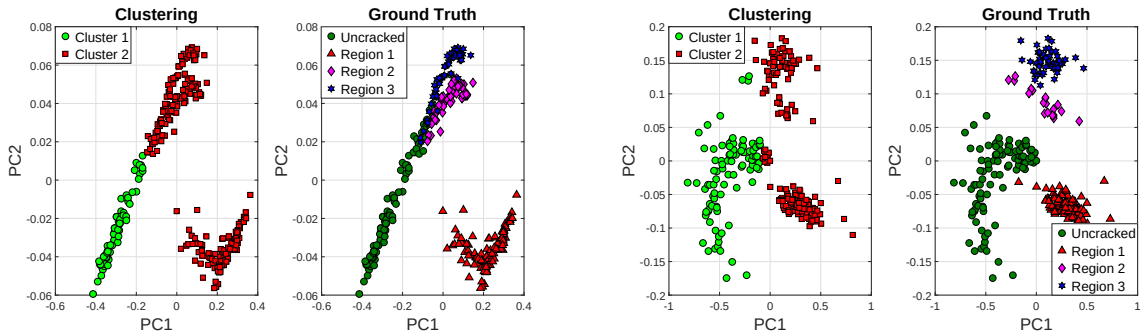
(a) Sensor at 0°

(b) Sensor at 60°

Figure 12: Results obtained from pipe B, when the actuator-sensor pair is away from the boundary. Damage is located at an angular position of 0°, in all the regions

difficult for the algorithm to classify damages in regions 2 and 3 as damaged situations. This is because, the damage signals from regions 2 and 3 are based on reflected signals obtained from damage.

Similarly, for a damage angular position of sixty degrees, as shown in figure 13, the response from the sensor at sixty degrees performs better than the one at zero degrees. We observe that for pipe B, beyond a damage offset of ninety degrees, the performance of the semi-supervised approach degrades to the point that it becomes unreliable for damage detection. However, the key take away from this discussion is that even with very few low quality piezo-sensors as transmitter and receiver, we can perform damage detection up to damage angular offsets of ninety degrees.



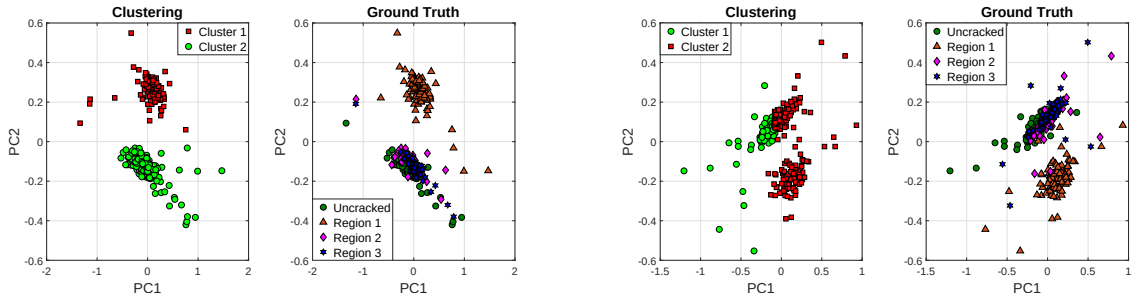
(a) Sensor at 0°

(b) Sensor at 60°

Figure 13: Results obtained from pipe B, when the actuator-sensor pair is away from the boundary. Damage is located at an angular position of 60° in all the regions

6.2.2. Close to the boundary

In this section we study the effects of a boundary on the performance of the proposed semi-supervised algorithm for damage detection.



(a) 0 degrees

(b) 60 degrees

Figure 14: Results for pipe B when the actuator-sensor pair is near the boundary. Here we show the results for sensor 3

As expected we observe a degradation in the performance of the proposed approach. This is due to the additional reflections from the boundary that corrupt the acquired signals at the sensors. In addition we also expect a greater energy dissipation of signals due to interaction with the supports. Due to additional reflection signals, damage responses from corresponding regions 2 and 3 are adversely affected as they are reflection based. We demonstrate that damage detection is still possible in region 1.

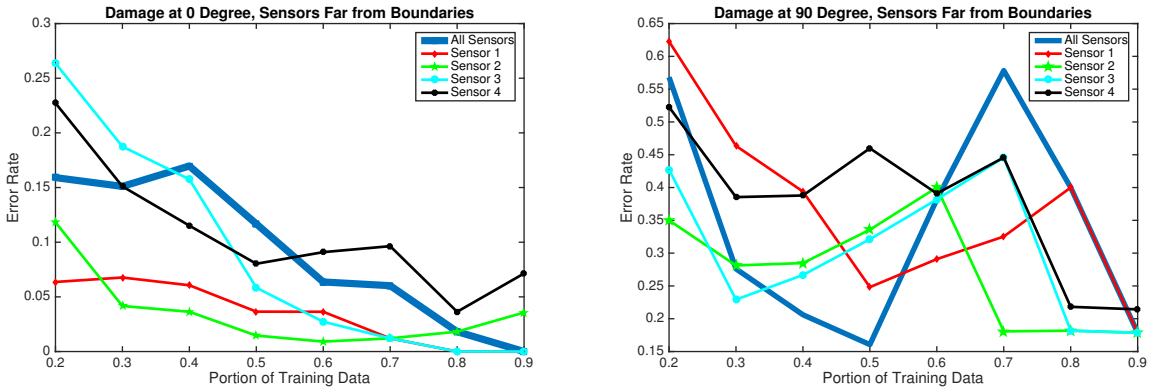
Figure 14 shows the results for damage angular locations of zero and sixty degrees. These results are based on a sensor angular offset of sixty degrees. For damage located at zero degrees, there is no distinction in between undamaged scenarios and damage located in regions 2 and 3. However, the proposed approach is able to detect damage when located in region 1.

7. Results: Supervised approach

In this section, we discuss the application of the proposed supervised approach for damage localization in this section. For this experiment, we use 1096 samples where each sample is a time series of length 2500 points. From these 1096 samples, 296 samples correspond to an undamaged pipe and the rest correspond to a damaged pipe where damages are located in 3 different regions. Our goal is to recognize whether or not a given sample signal corresponds to an undamaged or damaged pipe. Furthermore, for damaged pipes, we are interested in localizing the damage to a specific region on the pipe (as defined in figure 5). Therefore, we have a 4-class classification problem. For this problem, we use MLR as described in Section 2.3. One

major issue with this classification problem is that we only have 1096 samples while every sample has 2500 points. This leads to a high-dimensional classification problem that degrades the classifier performance. We use the PCA technique as described in Section 2.1 in order to decrease the dimensionality of every sample. In this set of experiments, we reduce the dimensionality of every sample from 2500 points to 150 principal components. We have 4 sensors located at different angles as described earlier. Figure 15 demonstrates the classification performance by MLR in two different scenarios. The *error rate* is defined as the ratio of number of false predictions to total number of test samples. In the first and second scenario, damage is located at 0 and 90 degree, respectively. Since our data is limited, we divide it into two different portions of training and test data. Figure 15 denotes the classification performance for different portions of training data. As we observe in Figure 15, when we use a larger portion of available data as training data, we obtain a better classification result on our test set; using 90% of the training data, the classification error rate drops to zero when the damage is located at 0°. Furthermore, our supervised approach performs better when damage is located at 0 degree compared to 90 degree. This is because as the offset of the damage location increases, the signal energy reaching the crack reduces further due to larger dissipation and hence the effects of the presence of a crack become obscured in the acquired signals.

Figure 16 shows the sensitivity of classifier performance to the sensor location. As we can see in this figure, when the sensor is close to boundaries, we have a degradation in classifier performance. The main reason for this phenomenon is signal reflections from pipe boundaries that act as a noise in original signal and hence, make classification more difficult.



(a) Damage at 0°

(b) Damage at 90°

Figure 15: Classification Performance of two scenarios: (a) Crack is located at 0°. (b) Crack is located at 90°. In each scenario, we get data from 4 sensors. These plots show classification performance for different sensors and portion of training data.

8. Conclusions

We propose a semi-supervised learning algorithm for damage detection and a supervised learning algorithm for damage localization for active sensing in pipes. The semi-supervised approach requires *a priori* label for just one undamaged data. It then uses a hierarchical clustering scheme for damage detection, efficiently separating damaged and undamaged cases using low-profile piezo-sensors up to damages located up to 90 degrees offset angles from the zenith line of the cylindrical pipe, where the actuators are located. We demonstrate the efficiency of the proposed approach for data acquired from both away from boundaries and close to them.

The supervised approach to damage localization is based on a multinomial logistic regression framework. The learner is trained using data acquired from a real pipe. Our experiments demonstrate that, given enough training data, our fully data-driven supervised approach accurately classifies the cracks into their originating regions. Therefore, this approach can be successfully used as a method for damage localization on complex

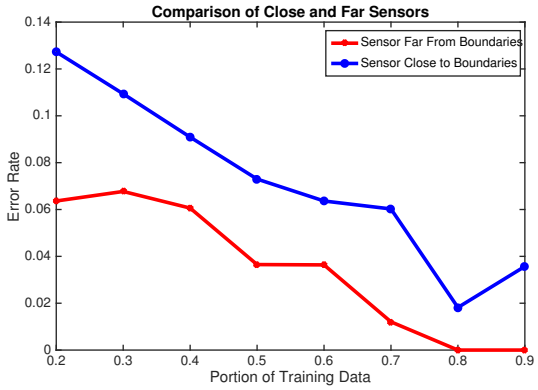


Figure 16: Effect of the distance of sensor from boundaries of pipe. As we can observe, when the sensor is farther from boundaries, we have a better classification performance.

structures such as pipes. We envision that our data-driven approach can be used for health monitoring of even more complex structural systems such as a network of pipes which we defer to a future work.

9. Acknowledgement

The authors gratefully acknowledge the funding from Texas Instruments (TI), TI-168 G84040/G83198, for this project. The authors thank Dr. Amardeep Sathyanarayana and Dr. Domingo Garcia at TI and Albert D. Neumann at Rice University for their help in setting up of the hardware for the experiments.

References

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

[2] A. Rytter. *Vibration basded inspection of civil engineering structures*. PhD thesis, Aolborg University, 1993.

[3] C.E. Restrepo, J.S. Simonoff, and Zimmerman R. Causes, cost consequences, and risk implications of accidents in US hazardous liquid pipeline infrastructure. *International Journal of Critical Infrastructure Protection*, 2(1-2):38–50, 2009.

[4] M.H. Park, I.S. Kim, and Y.K. Yoon. Ultrasonic inspection of long steel pipes using lamb waves. *NDT and E International*, 29(1):13–20, 1996.

[5] M.J.S. Lowe, D.N. Alleyne, and P. Cawley. Defect detection in pipes using guided waves. *Ultrasonics*, 36(1-5):147–154, 1998.

[6] Z. Liu and Y. Kleiner. State-of-the-Art Review of Technologies for Pipe Structural Health Monitoring. *IEEE Sensors Journal*, 12(6):1987–1992, 2012.

[7] J.L. Rose. Ultrasonic Guided Waves in Structural Health Monitoring. *Key Engineering Materials*, 270-273:14–21, 2004.

[8] J.L. Rose, Y. Cho, and M.J. Avioli. Next generation guided wave health monitoring for long range inspection of pipes. *Journal of Loss Prevention in the Process Industries*, 22(6):1010–1015, 2009.

[9] L. Yu, V. Giurgiutiu, J. Wang, and Y-J. Shin. Corrosion detection with piezoelectric wafer active sensors using pitch-catch waves and cross-time-frequency analysis. *Structural Health Monitoring*, 11(1):83–93, 2011.

[10] V. Giurgiutiu. *Structural Health Monitoring: with Piezoelectric Wafer Active Sensors*. Academic Press, 2008.

[11] H. Lee, J. Yang, and Sohn. H. Baseline-free pipeline monitoring using optical fiber-guied laser ultrasonics. *Structural Health Monitoring*, 11(6):684–695, 2012.

[12] W-B Na and Kundu T. Underwater Pipeline Inspection Using Guidedl Waves. *Transactions of the ASME*, 124:196–200, 2002.

[13] Pipeline Monitoring Systems Market by Technology (Ultrasonic, PIGs, Smart Ball and Magnetic Flux Leakage), Application (Leak Detection, Operating Efficiency and Pipeline Break Detection), End-Use Industry and Pipe Type - Global Forecast to 2026. `http://www.marketsandmarkets.com/Market-Reports/pipeline-monitoring-system-market-138709351.html`. Accessed: 2017-05-09.

[14] D.N. Alleyne, M.J.S. Lowe, and P. Cawley. The Reflection of Guided Waves from Circumferential Notches in Pipes. *Journal of Applied Mechanics*, 65(3):635–641, 1998.

[15] A. Demma, P. Cawley, M.J.S. Lowe, A.G. Roosenbrand, and B. Pavlakovic. The reflection of guided waves from notches in pipes: a guide for interpreting corrosion measurements. *NDT and E International*, 37(3):167–180, 2004.

[16] Y. Lu and J.E. Michaels. Numerical implementation of matching pursuit for the analysis of complex ultrasonic signals. *IEEE transactions on ultrasonics, ferroelectrics and frequency control*, 55(1):173–182, 2008.

[17] D.-A. Tibaduiza, M.-A. Torres-Arredondo, L.E. Mujica, J. Rodellar, and C.-P. Frintzen. A study of two unsupervised data-driven statistical methodologies for detecting and classifying damages in structural health monitoring. *Mechanical Systems and Signal Processing*, 41:467–484, 2013.

[18] D. Sen, A. Aghazadeh, A. Mousavi, S. Nagarajaiah, and R. Baraniuk. Sparsity based approaches for damage detection in plates. *Mechanical Systems and Signal Processing*, 117:333–346, 2019.

[19] R. Malladi, A. Dabak, and N.K. Murthy. Modelling ultrasound guided wave propagation for plate thickness measurement. In *Nondestructive Characterization for Composite Materials, Aerospace Engineering, Civil Infrastructure, and Homeland Security*, 2014.

[20] C. Liu, J.B. Harley, M. Berges, D.W. Greve, and I.J. Oppenheim. Robust ultrasonic damage detection under complex environmental conditions using singular value decomposition. *Ultrasonics*, 58:75–86, 2015.

[21] Y. Ying, J.H. Garrett Jr., J. Harley, I.J. Oppenheim, J. Shi, and L. Soibelman. Damage Detection in Pipes under Changing Environmental Conditions Using Embedded Piezoelectric Transducers and Pattern Recognition Techniques. *Journal of Pipeline Systems Engineering and Practice, ASCE*, 4(1):17–23, 2013.

[22] Y. Ying, J.H. Garrett Jr., I.J. Oppenheim, L. Soibelman, J. Harley, J. Shi, and Y. Jin. Toward Data-Driven Structural Health Monitoring: Application of Machine Learning and Signal Processing to Damage Detection. *Journal of Computing in Civil Engineering, ASCE*, 27(6):667–680, 2013.

[23] M. Eybpoosh, M. Berges, and H.-Y. Noh. Sparse representation of ultrasonic guided-waves for robust damage detection in pipelines under varying environmental and operational conditions. *Strcutural Control and Health Monitoring*, 23(2):369–391, 2016.

[24] T. Hastie, R. Tibshirani, and J. Friedman. *Elements of Statistical Learning, Data Mining, Inference and Prediction*. Springer, 2009.

[25] A. Raghavan and C.E.S. Cesnik. Review of Guided-wave Structural Health Monitoring. *The Shock and Vibration Digest*, 39(2):91–114, 2007.

[26] J. Li and J.L. Rose. Natural beam focusing of non-axisymmetric guided waves in large-diameter pipes. *Ultrasonics*, 44:35–45, 2006.

[27] E. Dehghan-Niri and S. Salamone. A multi-helical ultrasonic imaging approach for structural health monitoring of cylindrical structures. *Structural Health Monitoring*, 14(1):73–85, 2014.