DATA-DRIVEN ANALYSIS OF ULTRASONIC PRESSURE TUBE INSPECTION DATA

P. Zacharis, G. M. West, G. Dobie, T. Lardner and A. Gachagan
Department of Electronic and Electrical Engineering
University of Strathclyde
Royal College Building, Glasgow, G1 1XW
panagiotis.zacharis@strath.ac.uk; graeme.west@strath.ac.uk; gordon.dobie@strath.ac.uk;
timothy.lardner@strath.ac.uk; a.gachagan@strath.ac.uk

ABSTRACT

Pressure tubes are critical components of the CANDU reactors and other pressurized heavy water type reactors, as they contain the nuclear fuel and the coolant. Manufacturing flaws, as well as defects developed during the in-service operation, can lead to coolant leakage and can potentially damage the reactor. The current inspection process of these flaws is based on manually analyzing ultrasonic data received from multiple probes during planned, statutory outages. Recent advances on ultrasonic inspection tools enable the provision of high resolution data of significantly large volumes. This is highlighting the need for an efficient autonomous signal analysis process. Typically, the automation of ultrasonic inspection data analysis is approached by knowledge-based or supervised data-driven methods. This work proposes an unsupervised data-driven framework that requires no explicit rules, nor individually labeled signals. The framework follows a two-stage clustering procedure that utilizes the DBSCAN density-based clustering algorithm and aims to provide decision support for the assessment of potential defects in a robust and consistent way. Nevertheless, verified defect dimensions are essential in order to assess the results and train the framework for unseen defects. Initial results of the implementation are presented and discussed, with the method showing promise as a means of assessing ultrasonic inspection data.

Key Words: CANDU, pressure tube, ultrasonic inspection, unsupervised, machine learning

1   INTRODUCTION

Ultrasonic non-destructive testing is widely used for flaw detection and characterization in materials within the nuclear industry [1, 2]. However, the analysis of data obtained from ultrasonic inspection is time consuming and often requires manual assessment due to the complexity of distinguishing between noise and signals indicating defective areas. Hardware advancements in ultrasonic non-destructive technologies offer the ability to stream data in far greater volumes than in the past. This is driving a demand for automatic inspection data analysis frameworks that will allow faster assessment while maintaining a consistent level of results. Consequently, several automation techniques have been explored in different industries, using either knowledge-based or supervised data-driven approaches [3-5]. These approaches usually involve rules, assumptions and signal transformations that are specific to the underlying monitored system, creating system-specific procedures, or require pre-labeled individual signals.

Currently, there is limited research focusing on the unsupervised analysis of ultrasonic inspection data. Outlier analysis has been applied as a means to diagnose the structural damage of a variety of systems [6-7]. This approach is based on the assumption of an existing baseline distribution of 'healthy' signals and uses a pre-defined threshold of three standard deviations to determine the state of the examined signals. More recent work [8] utilizes a clustering algorithm and illustrated the feasibility of the

method on polymer specimen containing artificial defects. However, two crucial issues make this method unsuitable for our purpose. Firstly, is the requirement of pre-defined number of clusters (see Section 3.3.1). Secondly, the assumption of point-to-point consistency between the signals, a requirement for the applied pre-processing step, cannot be guaranteed by the current pressure tube inspection process.

In this work, we propose an unsupervised machine learning approach to group together signals derived from defective regions, and separate them from those derived from healthy regions of a component under inspection. The process requires no individually labeled signals, nor a predefined normal state. It follows a two-stage clustering procedure that takes into account both signal features and position features, enabling noise elimination. The extracted signal features are simple statistical properties of the signals that require no wave-transformations. Furthermore, the number of features is minimal to avoid the curse-of-dimensionality and time-consuming feature-space transformations that increase the number of the assumptions. Enabling this procedure is a density-based clustering algorithm (DBSCAN) which offers advantages for this application, such as the detection of arbitrarily shaped clusters, no requirement to predefine the number of clusters, and outlier detection.

This paper describes the application of the machine learning approach to inspection data of CANDU reactor pressure tubes. The datasets consist of ultrasonic signals obtained by a 20MHz normal incidence pulse echo probe and two pairs of axial and circumferential pitch-catch probes from areas containing defects. Their location and sizing have been verified manually by independent analysts, providing a valuable source for the algorithm's assessment stage. Initial results of the implementation are presented and discussed, with the method showing promise as a means of identifying pressure tube defects.

2   CANDU REACTORS

2.1 CANDU Design

CANDU is a type of pressurized heavy water reactor (PHWR) that uses natural uranium as fuel and heavy water (deuterium oxide D_2O) as both coolant and moderator. The core structural components consist of low neutron absorbing materials (zirconium alloys) and its design enables high neutron economy which is critical for achieving fission through a sustained chain reaction, given that natural uranium has low fissile content. The core consists of a low-pressure steel tank, called the calandria, which contains the moderator and around 400 fuel channels that run across the length of the calandria tank. Each fuel channel (Fig. 1) consists of a thin zircaloy tube (calandria tube) that contains another, thicker, zircaloy tube, called pressure tube. The two tubes are separated by the annulus garter-spring spacers that create a gap allowing the flow of an insulating gas, called annulus gas. The fuel bundles are located inside the pressure rubes, and they are cooled by the hot (~300 ^{o}C) pressured (~10 MPa) heavy water coolant that runs through the pressure tubes. This configuration enables the insulation of the unpressurised cool (~70 ^{o}C) moderator from the hot pressurized coolant.
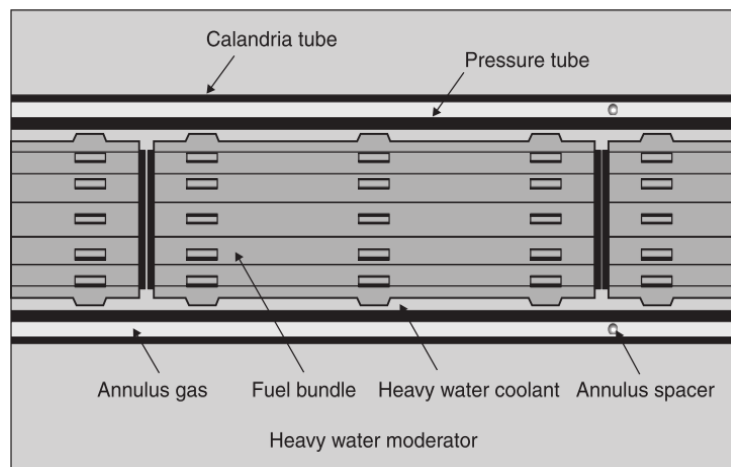

Figure 1. CANDU reactor fuel channel [10].

2.2 Pressure Tubes

Pressure tubes are one of the critical components of pressurized heavy water reactors currently operating in Argentina, Canada, India, Pakistan, the Republic of Korea, and Romania. During their 30 year design life they are required to operate reliably in an extremely harsh environment of high pressure, temperature and neutron flux. Although the extensive quality controls and monitoring of the pressure tubes have ensured a good overall performance [9], pressure tubes are subjected to some degradation mechanisms facilitated by the severe conditions, which could cause a pressure tube failure.

One of the principal degradation mechanisms that pressure tubes experience is concentration of hydrogen/deuterium at large tensile stresses within the tube. This can precipitate hydrides that initiate a process called Delayed Hydride Cracking (DHC) that can cause coolant leakage potentially damaging the reactor. Therefore there is a need to closely monitor any pressure tube tensile stress concentrations, such as flaws and defects created during tube manufacture, installation, commissioning or operation. Examples of such flaws include refueling scratches, fuel fretting, crevice corrosion and debris fretting.

2.3 Inspection Setup

The inspection of CANDU pressure tubes is currently mainly performed during planned, statutory outages. This paper is focuses on data obtained by a multiple ultrasonic probe tool known as CIGAR (Gauging Apparatus for Reactors). The tool consists of two normal incidence probes (NB), operating at frequencies of 10MHz and 20MHz, and two pairs of axial and circumferential pitch-catch probes (APC and CPC respectively) [11]. Fig. 2 illustrates a simplified configuration of a 20MHz NB probe and the two pitch-catch pairs that effectively provide a multi-view of a defect.

The process of analyzing these datasets of inspection data to provide defect assessment is currently being carried manually by expert analysts. This constitutes a time consuming and costly operation as the reactor cannot be restarted until the end of the assessment process.

Currently, a new more advanced tool known as BRANDE (BRIMS Advanced Non-Destructive Examination) is being deployed which is able to provide high resolution data of significantly larger volume. This is highlighting the need for an autonomous signal analysis process. The next section describes the proposed data-driven framework that aims to provide decision support for the assessment of potential defects within the CANDU pressure tubes, in a robust and consistent way.
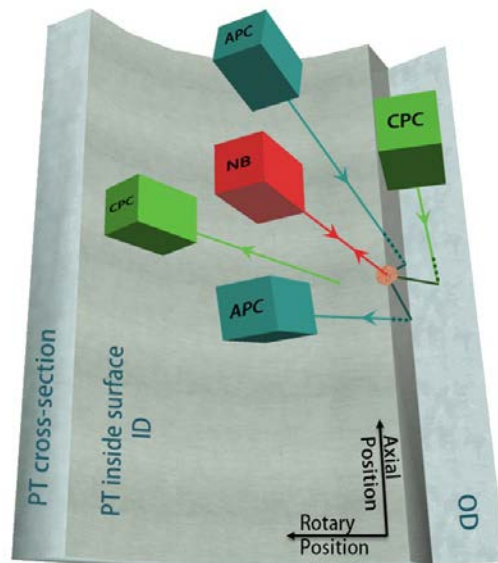

Figure 2. Orientation of inspection probes.

3 PROPOSED FRAMEWORK

3.1 Ultrasonic Inspection Data

This research focuses on datasets containing the inspection ultrasonic waveform data obtained by the 20MHz NB, the APC and the CPC probes of the CIGAR tool from sections of the inside surface of several pressure tubes. The tool scans the area of interest in axial and rotary increments effectively providing a set of three waveforms for each of the individual positions. The recorded waveforms consist of sampled amplitudes and are presented in the form of A-Scans (Fig. 3).
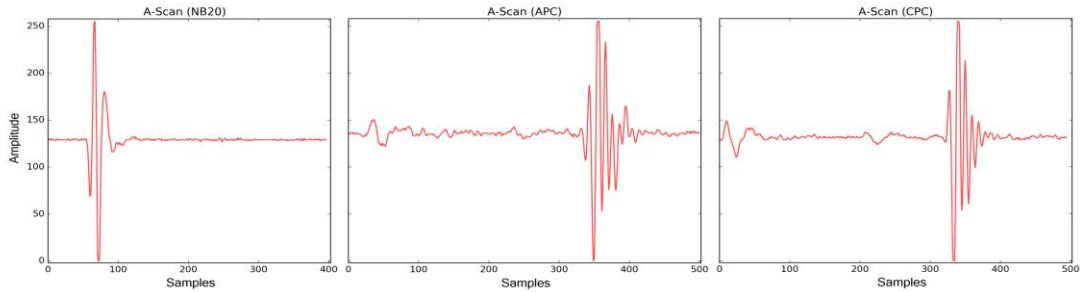
Figure 3. Examples of A-Scans obtained by different probes from the same position.

Each dataset is obtained from separate areas of interest that are known to contain defects of debris fretting type. However, the available A-Scans are unlabeled, i.e. the individual signals are not accompanied by the information of whether their location is within or outside the often irregularly shaped defective area. Instead, the information that will be used for assessing the performance of the following analysis is the verified dimensions and starting position of the defects.

3.2 Feature Selection

As the aim of this research is to identify and isolate groups of signals, it is useful to condense the information contained in each signal (average of 490 data points of amplitude-time pairs) into small subsets of representative features which act as the input to the learning process. Driven by the requirement to provide a general system that is able to adapt to the diverse surface pressure tube conditions and the irregularly shaped defects, the features are selected based on a set of characteristics that aims to facilitate the formation of a non-system-specific procedure.

The basis of the proposed selection is that the features should be descriptive characteristics of the individual signals that do not require explicit threshold-based assumptions about the signal. Therefore, descriptive statistical measures that quantitatively summarize the entire signal were chosen instead of gating methods or methods based on detection of explicitly defined signal characteristics. Furthermore, the signals are not filtered through signal transformation procedures, such as Fourier Transform (FT) and Discrete Wavelet Transform (DWT). FT, besides increasing the computational effort, does not provide time-related information. Although DWT provides temporal resolution, it would increase the complexity
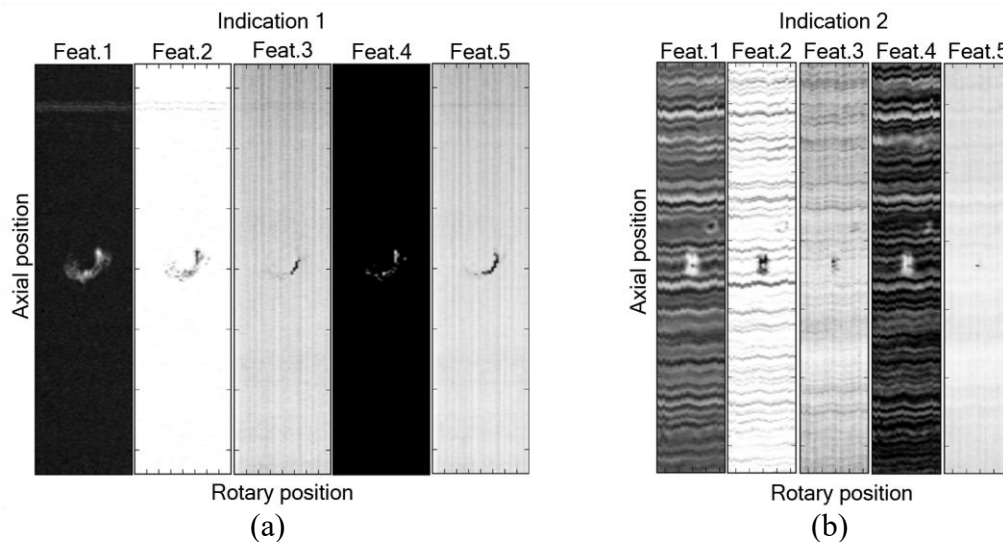

Figure 4. Examples of C-Scans generated for two areas containing defects, utilising the values of individual features.

of the system as it is sensitive to the choice of wavelet type as well as the length of the wavelet filter [12].

The extracted signal features are: the variance, the maximum value, the minimum value, and the index positions of the first instances of maxima and minima within the signal. These features, although simple, are showing promising capabilities of carrying information about the condition of the pressure tube surface and potential defects. A way of illustrating this information is a two-dimensional top view presentation (C-Scan) of the dataset where each pixel represents the physical position of a signal, and its intensity represents the value of the corresponding signal's feature (e.g. Fig. 4).

Fig. 4a illustrates an example where the condition of the inside surface of the pressure tube enables each feature to produce a clear view of the location and shape of the defect. However, this is not the general case as the presence of various factors, such as foreign material or/and accumulated oxide on the surface of pressure tubes and within the defect, can cause additional signal scattering and reflections creating noisy patterns that are more challenging to be addressed by the features (e.g. Fig. 4b). The random behavior of these factors poses a problem for modelling and predicting the degree to which they affect the performance of the features on each unique pressure tube environment. Therefore, the proposed methodology utilizes the full set of equally weighted features to represent the signals in the unsupervised learning process.

3.3 Unsupervised Signal Analysis

3.3.1 DBSCAN clustering algorithm

The problem of grouping together unlabeled ultrasonic signals is being approached by an unsupervised clustering machine learning algorithm called DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [13]. DBSCAN is based on the idea that data points in dense regions will form a cluster while data points from different clusters will be separated by low density regions. This approach differs from other clustering algorithms, such as k-means [14] and mean shift [15] where the clusters are defined by their centers. This causes them to underperform when processing non-spherical shaped clusters. An illustration of its capabilities, versus k-means and mean shift, is shown on Fig. 5 which illustrates their inability to handle arbitrarily shaped clusters. It also shows another disadvantage of k-means, which is the predefined number of clusters. Since the proposed application is aimed at handling inspection signals, the algorithm should be able to determine the different (a priori unknown) number of groups and not forcing the data to fit the assumptions about the dataset. The three advantages of DBSCAN which are critical for our application are:

Figure 5. Comparison of k-means, Mean Shift and DBSCAN [16].

- Detection of arbitrarily shaped clusters

- No assumption about the number of clusters

- Detection of noise and outliers

DBSCAN requires two input parameters (*eps*: Maximum radius of the neighborhood, *MinPts*: Minimum number of points in the eps-neighborhood of a point) and operates by iteratively distinguishing each data point as one of the following types (Fig. 6):
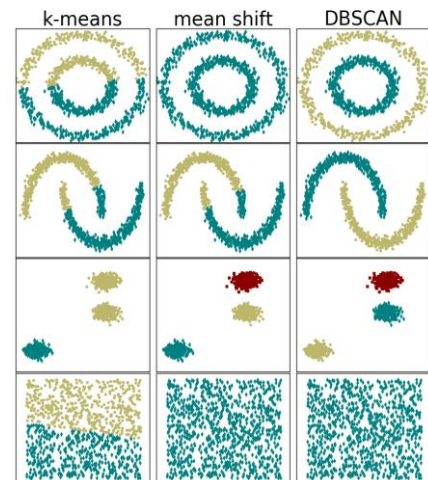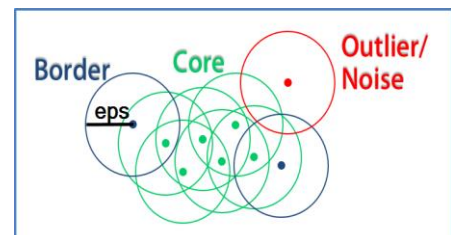

Figure 6. Example of DBSCAN cluster with parameters MinPts: 3 and eps: radius.

- Core point: point with a dense neighborhood

- Border point: point that belongs to a cluster but its neighborhood is not dense

- Noise point: point that does not belong to any cluster

3.3.2   Feature standardization

After the feature extraction step, each signal is represented by its corresponding set of feature values. This enables greater computational efficiency during the subsequent clustering procedure due to the inherent reduction of the search space. Preceding the clustering procedure is a pre-processing step which reformats the features onto the same scale (feature scaling). This is required by algorithms which use distance metrics (such as the Euclidean distance) as the rescaled values of the features will ensure that the computed distances are not dominated by features with intrinsically large numeric values. Two common approaches to scaling are: *normalization* and *standardization*. Normalization is often approached as a case of min-max scaling where the feature values are rescaled to a range of [0, 1]. This work uses the standardization approach which rescales the feature values so that they are re-centered around zero (zero mean) with standard deviation 1. The calculation of the new standardized value $x_{std}^{(i)}$ of a sample $x^{(i)}$ is expressed by Equation 1:

$$x_{std}^{(i)} = \frac{x^{(i)} - \bar{x}}{\sigma} \tag{1}$$

where $\bar{x}$ is the sample mean of the particular feature, and σ is the corresponding standard deviation.

3.3.3   Two-stage clustering

After pre-processing, the standardized features are processed by the first stage of the clustering procedure using the DBSCAN algorithm. This step aims at reducing the size of the sample by eliminating as many as possible 'healthy' signals, and isolate signals of ambiguous nature. The output of this clustering is a label attached to each unique feature set, indicating the class/group that it belongs. These labels are also mapped back to any samples that did not participate in the calculations because they shared one of the unique set of feature values. Given the unbalanced nature of the datasets, i.e. the healthy section of the tested area is larger than the defective section, an assumption is made that the group with the larger number of members belongs to the healthy section, thus the remaining groups contain potential defective areas or noise. Therefore, only the smaller groups of samples will proceed to the next clustering stage.

The samples that proceed at the second stage of the clustering procedure are being represented only by their location properties, i.e. *axial position* and *rotary position*, and are all considered as potential 'defective' signals. The dual purpose of this stage is to:

- Eliminate outliers/noise by identifying samples that are not members of the 'healthy' group and are located in a sparse neighborhood.

- Group together samples that are not members of the 'healthy' group and are located in a dense neighborhood.

This step is essential for formulating the final group comprising of samples which are members of the defect. It provides the means to group together signals that although obtained from a defective region, might not be initially grouped together; a consequence of the diverse characteristics that constitute a defect and their different impact on the reflected signals. Therefore, the location-based clustering is aiming at re-grouping those signals together, utilizing the fact they should be located in close proximity.
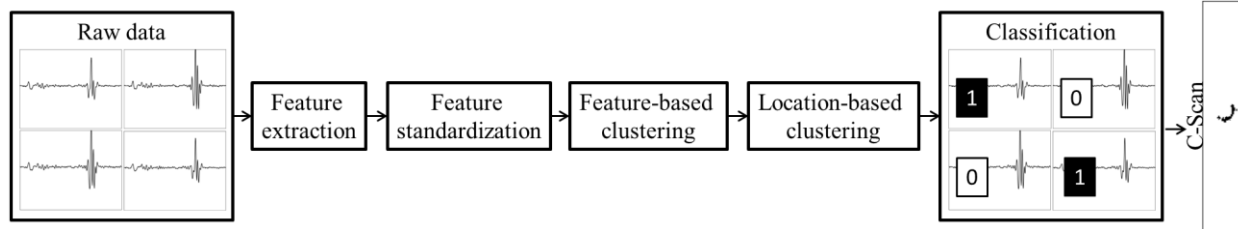
Figure 7. Flowchart of the clustering procedure.

The output of this location-based clustering is the updated signal labels: Samples that belong to a group are classified as part of the defective region, and samples that were classified as noise/outliers are discarded as being members of the healthy region. Fig. 7 provides a high-level flowchart that summarizes the steps taken by the clustering procedure. This process can be applied to the different datasets obtained by each of the ultrasonic probes (APC, CPC and 20 MHz NB).

4 RESULTS

Fig. 8 presents examples of C-Scans after applying the signal clustering procedure to three different Indications (areas that contain defects) for each of the datasets obtained by the different ultrasonic probes. Each pixel represents the label assigned to the corresponding A-Scan: 'white' represents signals classified as 'healthy' and the darker colors represent 'defective' signals. However, there are cases where locations of the healthy section are indicated as 'defective' (e.g. the grey areas of Fig. 8a APC and Fig. 8b APC). This is a result of the different conditions of the pressure tube and the area surrounding the defect, affecting heavily the propagating signals. However, the diversity of the probe configuration should enable future work to eliminate the 'noisy' clusters by fusing the results of the different probes.
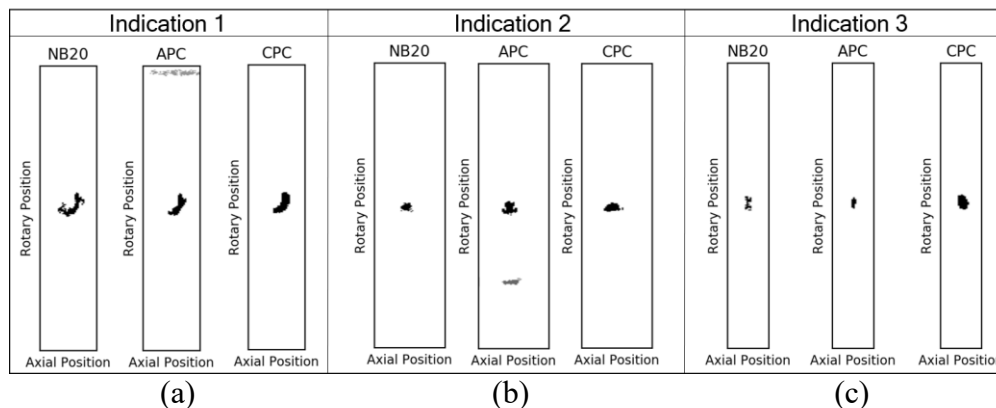

Figure 8. Examples of C-Scans resulted from the signal clustering procedure.

Currently, the algorithm requires manual tuning of a clustering parameter (eps). To explore the sensitivity of the system to the values of the parameter, the algorithm was applied to 37 unique indications and the results were compared to the provided verified defect dimensions. Fig. 9 presents how many (percentage) of the available indications can each of the eps values measure their width in absolute difference less than 25%, for each of the ultrasonic probes.

The results indicate that no single parameter value can satisfy each of the unique indications. The amount of information about the defect that is available for extraction from a dataset depends on the condition of the pressure tube (foreign material or corrosion), the characteristics of the defect, and the

behavior of each inspection probe under the given former conditions. Therefore, there is a need for training the algorithm to automatically identify the most appropriate clustering parameter for each case.
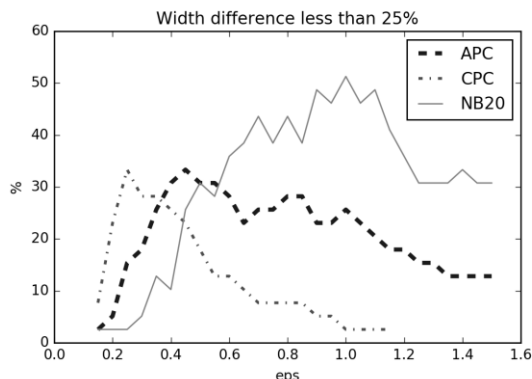

Figure 9. Application of different eps values to 37 indications. Percentage of achieved absolute width difference below 25%.

5   FURTHER WORK

The most important step for refining this work is the implementation of a training procedure that will create a predictive model for providing the appropriate eps parameter value given some characteristics of the tested area. Initially this will require a training dataset containing extracted attributes of known processed datasets, along with the eps value that produced the best result on each dataset.

However, one factor is limiting the current work to approach the training procedure. First, there is a need to define what makes a result better than another. By manually assessing the current produced results from various eps values, it has been identified that there are cases where the selection of the best result can be an ambiguous process. An example of this would be a result that matches the verified defect dimensions but also returns large amounts of noise, or a noiseless result that underestimates the defect dimensions. It is crucial to standardize the process of selecting the best eps for the known verified defects, as they constitute the labels that will be used for the training process of the predictive selection model.

Furthermore, although the current work developed a framework that can be applied to datasets captured by ultrasonic probes of different configurations, it does not utilize the extracted information in a unified way. Fusing results obtained from different probes could potentially provide a more informed estimation of the defects' dimensions and shape, as well as elimination of noisy clusters. The fusion of these results could be approached by treating each isolated probe model as a classifier and form an ensemble learning scheme [17].

6   CONCLUSION

Defects on the inside surface of CANDU pressure tubes facilitate DHC which can lead to coolant leakage and can potentially damage the reactor. This work has introduced a new approach for identifying these defects through data-driven analysis of the available unlabeled pressure tube ultrasonic inspection signals. The framework uses simple descriptive statistics as features and a two-stage unsupervised density-based clustering procedure and it was tested on real datasets obtained from the inspection of several pressure tubes. Initial results were discussed, with the method showing promise as a means to detect and measure defects. Furthermore, this paper discussed the limitations of the proposed system and proposed future steps to improve its performance and eliminate the need for manually selecting a clustering parameter.

7   ACKNOWLEDGEMENTS

8   REFERENCES

1. G. Dobmann, N. Meyendorf and E. Schneider, "Nondestructive characterization of materials A growing demand for describing damage and service-life-relevant aging processes in plant components", *Nuclear Engineering and Design*, **171(1-3)**, pp.95-112 (1997).

2. G. Sposito, C. Ward, P. Cawley, P. B. Nagy, and C. Scruby, "A review of non-destructive techniques for the detection of creep damage in power plant steels", *NDT & E International*, **43(7)**, pp.555-567 (2010).

3. S. Iyer, S. K. Sinha, B. R. Tittmann, and M. Pedrick, "Ultrasonic signal processing methods for detection of defects in concrete pipes", *Automation in Construction*, **22**, pp.135-148 (2012).

4. F.W. Margrave, K. Rigas, D.A. Bradley, and P. Barrowcliffe, "The use of neural networks in ultrasonic flaw detection", *Measurement*, **25(2)**, pp.143-154 (1999).

5. T.D'Orazio, M. Leo, A. Distante, and G. Cavaccini, "Automatic ultrasonic inspection for internal defect detection in composite materials", *NDT & E International*, **41(3)**, pp.145-154 (2008).

6. P. Rizzo, M. Cammarate, D. Dutta, H. Sohn, and K. Harries, "Unsupervised learning algorithm for fatigue crack detection in waveguides", *Smart Materials and Structures*, **18(2)**, 025016 (2009).

7. X. Zhu and P. Rizzo, "Sensor array for the health monitoring of truss structures by means of guided ultrasonic waves", *Journal of Civil Structural Health Monitoring*, **4(3)**, pp.221-234 (2014).

8. R. You, Y. Yao, J. Shi, K. Zheng, and K. H. Wang, "Feature-selective clustering for ultrasonic-based automatic defect detection in FRP structures", *Chemometrics and Intelligent Laboratory Systems*, **157**, pp.35-42, (2016).

9. IAEA, "Update of operating experience with cold-worked Zr-2.5%Nb pressure tubes in CANDU reactors", http://www.iaea.org/inis/collection/NCLCollectionStore/_Public/35/055/35055669.pdf (1991). Accessed February 1, 2017.

10. I. Crossland, *Nuclear Fuel Cycle Science and Engineering*, Woodhead Publishing Limited, Cambridge, UK (2012).

11. M. Trelinski, "Inspection of CANDU Reactor Pressure Tubes Using Ultrasonics", *17th World Conference on Nondestructive Testing*, China, 25-28 October, (2008).

12. Z. Zhang, Q. K. Telesford, C. Giusti, K. O. Lim, and D. Bassett "Choosing Wavelet Methods, Filters, and Lengths for Functional Brain Network Construction", *PLoS ONE*, **11(6)**, e0157243. http://doi.org/10.1371/journal.pone.0157243 (2016).

13. M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise", *KDD-96*, USA, (1996).

14. S. P. Lloyd, "Least Squares Quantization in PCM", *IEEE Transactions on Information Theory*, **22(2)**, pp.129-137 (1982).

15. Y. Cheng, "Mean shift, mode seeking, and clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **17(8)**, PP.790-799 (1995).

16. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot,

and E. Duchesnay, "Scikit-learn: Machine Learning in Python", *Journal of Machine Learning Research*, **12**, pp.2825-2830 (2011).

17. D. Parikh and R. Polikar, "An Ensemble-Based Incremental Learning Approach to Data Fusion", *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **37(2)**, pp.437-450 (2007).