Practical Guided Wave SHM of Pipes – Processing Multiple Data Sets to Give Reliable Defect Detection

Chang LIU 1, Peter CAWLEY 1

1 Dept Mechanical Engineering, Imperial College London, London SW7 2AZ, UK

Abstract
Historically, most guided wave testing has been done on a one-off test (NDT) basis. However, permanently installed systems are now commonly used, giving the opportunity to move from NDT to SHM. In conventional SHM, damage is detected by subtracting a measurement from a baseline record, after compensating for any temperature difference between the tests. However, the temperature compensation methods cannot perfectly remove the benign variations produced by complex environmental and operational conditions (EOCs), leaving residual noise that can mask the damage signal. Component analysis methods such as singular value decomposition (SVD) and Independent Component Analysis (ICA) operating on multiple data sets have the potential to improve the sensitivity and reliability of defect detection. In this paper, we evaluate the SHM performance on a synthetic dataset that contains the superposition of experimental guided wave records collected on a test pipe during temperature cycling, and computationally generated damage signals at various locations. This synthesis process enables us to investigate the performance of the methods under different EOC and damage conditions. We compute the receiver operating characteristic (ROC) curves that plot the probability of detection against the probability of false alarm for different damage growth rates and different frequency of measurements using the SVD, ICA, and the conventional baseline-subtraction residual methods. The results presented here show the methodology using the conventional baseline-subtraction method; the conference presentation will show the comparison with SVD and ICA. The results show that the sensitivity of the SHM processing is an order of magnitude better than that typically obtained in one-off guided wave NDT, and defects can much more reliably be detected at features such as welds. The ROC curves can efficiently be derived for an installed SHM system, so enabling the probability of detection for a given size of defect to be computed, together with the corresponding false call rate; the benefits of taking data more frequently can also be evaluated. This means that an installed system can be tuned to give the required performance and this performance can be validated, which promises to be very valuable in developing safety cases.

Keywords: PIMS, Guided wave, ROC curve, Singular value decomposition, Independent component analysis

1. INTRODUCTION
Guided wave testing has been widely used to detect damage on structures. Recent advances in sensing and computation technology make it attractive to use permanently installed ultrasonic transducers to monitor the integrity of structures, which can potentially improve the reliability and reduce the operating cost associated with regular inspections [1], [2].
In practical implementation of a long term monitoring system, damage is usually detected by comparing measurements with baseline records and seeking changes that represent defect signatures. The comparison can be done by subtracting the baseline from the current measurement, by calculating the cross-correlation between the measurements and the baseline [3], or by more advanced data-driven methods such as wavelet transform [4] or component analysis [5], [6]. If damage were the sole source of change in the duration of the monitoring, such a comparison would accurately reveal the progression of damage. However, guided wave records in long term monitoring are often inevitably affected by various environmental and operational conditions (EOCs), which degrade the performance of the damage detection [7], [8]. Therefore, it becomes critical to predict the performance of damage detection schemes under practical EOCs, to guide the practical implementation of such an evaluation.

One way to conduct such an evaluation is by conducting laboratory experiments under varying EOCs and physically growing defects on the test specimens. However, this approach can be prohibitively expensive, especially when the damage needs to be grown progressively and in various EOCs.

Another approach is to simulate the received signal – modern computational resources mean we can reliably predict the signature produced by damage, even when it has complex shape. However, reliable prediction of signal changes due to environmental and other variability is much more difficult. On the other hand, obtaining experimental data with environmental variation in an undamaged structure is easy. Therefore, we propose a methodology of measuring data over multiple environmental cycles on an undamaged structure and synthetically adding damage to signals.

This approach enables us to add damage at different locations with different growth patterns, and to easily investigate other practical parameters such as the degree of EOCs, damage severity, frequency of readings, etc. Another advantage is that by repetitively randomly selecting from the records that are collected in specific EOC ranges, we can create a sufficient number of data sets to statistically evaluate the performance of the damage detection in these practical settings. The statistical measure we choose to assess the damage detection is the receiver operating characteristic (ROC) curve, which is used in statistics to illustrate the performance of a binary classifier [9], and has been also widely adopted in NDE/SHM [4], [10].

In the following sections, we first illustrate our proposed methodology that computes the ROC of damage detection schemes on experimental datasets synthesized with artificial damage, then demonstrate an example implementation with experimental records taken on a pipe specimen under temperature cycles. We run the damage detection on the synthetic datasets and summarize the ROC curves obtained. We also discuss examples using the ROC curves to evaluate the damage detection performance in practical settings.

2. METHODOLOGY

[Figure 1] shows the flow diagram of the proposed methodology. We first collect experimental data on an intact structure under various environmental and operational conditions (EOCs) that are likely to occur in its operation. Such data collection can be done in a relatively short time by physically applying EOC variations. The experimental data then serves as the
undamaged baselines on which artificial damage is to be superposed.

We then specify the EOCs of interest and randomly select from experimental records collected under such conditions. For each test scenario, we repeat the randomization process multiple times to generate multiple synthetic datasets. The repetitive random sampling helps us avoid statistical outliers and get an accurate evaluation of the damage detection. We then superpose artificial damage signals to create the synthetic datasets. These datasets resemble what we would obtain from practical long term monitoring if there was a growing defect, without actually damaging the specimen.

![Process flow diagram of the proposed evaluation methodology](image)

Figure 1: Process flow diagram of the proposed evaluation methodology, which calculates ROC on synthetic datasets generated with experimental undamaged records and artificial damage signals. Shaded blocks represent data, blank blocks represent process/computation, and dashed blocks represent the pre-specified parameters.

Experimentally creating progressive damage on a large structural specimen can be prohibitively expensive, especially when the damage needs to be tested in various EOCs. However, it is relatively easy to predict the progression of the reflections from damage as it grows, using reflection coefficients of the defects predicted using methods described in [11], [12].

We then apply damage detection schemes on the synthetic dataset we created using experimental signals. The general implementation of such schemes includes temperature compensation, feature extraction, and damage classification. We first compensated the temperature change using the scale transform [13], an efficient stretch-based temperature compensation method. We then subtract the baseline signal from temperature compensated signals. The residual signals are then match filtered with the excitation pulse to improve the signal-to-noise ratio.

We need to statistically assess the effectiveness of the damage feature at different EOC and damage conditions. To achieve that, we create a large number of synthetic datasets by repetitive random sampling from undamaged experimental data, and compute the ROC curve of the damage feature.
In statistics, ROC curves show the true positive rate (TPR) against the false positive rate (FPR), in various threshold values. In the field of NDE/SHM [4], [10], ROC is also often used to illustrate the performance of damage detection, in which the two axes are termed the probability of detection (POD) and the probability of false alarm (PFA), both in the range of 0 to 1. The perfect classifier yields a curve that goes through point (0,1), indicating no false alarms and perfect (100%) detection, while a random guess yields a ROC curve follows the 45° diagonal line.

Figure 2: A Guided Ultrasonics, Ltd. permanent monitoring transducer ring was bonded to an 8" Sch.40 steel pipe specimen to excite and receive ultrasound records. A heating element was inserted into the pipe to create temperature variations.

2.1 Example implementation

Figure 2 shows a schematic drawing of the instrumented pipe specimen: a 6m-long NPS8 Schedule 40 steel pipe segment. A Guided Ultrasonics, Ltd. Permanently installed monitoring (gPIMS) transducer ring was bonded to the pipe, 2m from one of the cut ends. The transducer ring excites an 8-cycle Hanning-windowed tone-burst along the forward direction as labelled in Figure 2 and receives the reflected and scattered ultrasonic waves from that direction. Figure 3 shows a typical T(0,1) mode reflection, whose amplitude is normalized to the main reflector ± the first cut end reflection at 4m. We plot the features up to 12m along the pipe including the 2nd cut end reflection at 10m, and two smaller benign features at 6m and 8m, which are likely to be caused by the reflections from the backward direction that the direction control processing has not completely cancelled. In this example implementation, we use these smaller benign features to simulate other structural features, such as a weld or a bend.

We heated the pipe up to 90°C with a heating element inserted into the pipe, then slowly cooled it down to room temperature, during which we took ultrasonic records every 15 minutes. We repeated the temperature cycles multiple times. On average, each cooling phase lasts for 10.5 hours and consists of 42 records. As the temperature varies between 30-90°C the stretching factor varies roughly from 0.999 to 1.006, with respect to the baseline collected at 40°C.

In this example, we simulate the damage as a reflection, whose amplitude grows linearly from zero to 1% of the cut end reflection. This simplification omitted the frequency dependent scattering and mode conversion that happens at a real defect, but serves our purpose for the illustration of the ROC methodology.

We first estimate the temperature effect by calculating the stretching factors between each of the randomly selected undamaged records and the baseline. We then calculate the time delays of the reflection from a chosen damage location if affected by these stretching factors. We
delay and scale the reflections (an 8-cycle Hanning-windowed tone burst in our example), and then superpose them onto the corresponding undamaged signals.

Figure 3: Example experimental ultrasonic record collected from the experimental setup shown in Figure 2.

Figure 4(a) plots the temperature compensated, baseline subtracted residual signal in solid line, with its maximum amplitude normalized to unity; the dashed line shows the true damage location. To calculate the ROC, we vary the threshold (dot-dashed line) from 0 to 1, and classify all the values above the threshold as positives. The positives that fall within the true damage region are labeled as true positives or detection (solid circles), and the positives elsewhere are labeled as false alarms (hollow circles). At every threshold, we calculate the probability of detection and false alarm rate, which becomes one point on the ROC curve. The collection of the points becomes the ROC curve for this test scenario.

Instead of scalar values, our damage feature is a waveform that has a spatial length, producing a spatial ambiguity of the damage location. Assuming a damage reflection as a Hanning-windowed sinusoid that peaks at 1 and tails at 0, if we sought to ‘hit’ every point in the window length, we need a threshold that is smaller than the smallest point in the windowed sinusoid, which would be even smaller than the noise and create unnecessary false alarms. However, in industrial practice, we would classify a damage detector to be perfect if a sufficiently large portion of the damage feature is substantially larger than the noise. We address this problem by introducing a ‘tolerance level’ when we determine the ‘true damage region’. For example, a -6dB tolerance level corresponds to 50% of the full window. Once all points in this region are labeled as positive, the POD becomes unity. A lower tolerance level means that we need to hit more points in a larger region around the damage location to bring POD to unity.

Figure 4: ROC curve on example experimental residual. (a) Binary classification on example residual signal. (b) The corresponding ROC curves with different tolerance levels.
In Figure 4(b), we plot the ROC curves corresponding to different tolerance levels from -1 to -40dB. As the tolerance level goes down, the less tolerant we are to noise, and the more false alarms we have at any POD level. We propose using a -6dB tolerance level to be consistent with industrial common practice.

2.2 Results

We generate synthetic datasets with combinations of parameters as shown in Error! Reference source not found.. Each synthetic dataset is generated for one set of temperature range, damage size and number of readings; for each combination of parameters, 500 synthetic datasets are generated with repetitive randomly selected undamaged experimental records. To give the reader a sense of the magnitude of these datasets, the combination of the three parameters yields in a total of 2.6 million synthetic records to be generated.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>[Unit] Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature range</td>
<td>2, 10, 30, 60</td>
<td>[°C] temperature difference</td>
</tr>
<tr>
<td>Damage size</td>
<td>0.1, 0.3, 0.5, 0.7, 1.0</td>
<td>[%] amplitude of main reflector</td>
</tr>
<tr>
<td>No. of readings</td>
<td>10, 30, 50, 70, 100</td>
<td>Number of readings in a synthetic dataset</td>
</tr>
</tbody>
</table>

Table 1: Various parameters used in the generation of the synthetic data sets.

Combining the synthetic datasets we created, we can investigate different practical scenarios of interest with different levels of damage under different EOCs. We show one representative scenario as an example, in which we evaluate the ROC over different temperature ranges and damage sizes. We took synthetic datasets each comprising 100 readings and plotted the ROC for each combination of temperature range and damage size. Figure 5 (a-d) show ROCs corresponding to, respectively, 2, 10, 30 and 60 °C temperature range. The curves in each plot correspond to different damage sizes. Because we randomly selected the undamaged signals in each test scenario 500 times, we obtain 500 ROCs. In Figure 5, we plot the median ROC of the 500 runs, while the error bars indicate the interquartile of PODs at specific PFAs.

By comparing the five ROC curves in each plot, we can quantitatively illustrate how damage detection performance improves as the damage size becomes larger. By comparing the ROC curves from Figure 5 (a-d), we can see how performance deteriorates as the temperature range increases for a specific defect size.

We can then use the ROCs in Figure 5 to evaluate the performance of the baseline subtraction residual as a damage detector in various practical settings. For example, the dark solid line in Figure 5(b) represents the ROC for the scenario where temperature varies in a 10 °C range, and the damage ramps up to 1% of the cut end reflection. If we seek to detect the damage with at least a 90% POD, there would be a false alarm rate of 25%. Likewise, we can read from the same ROC curve that if we want to limit the false alarm rate to be below 20%, the corresponding detection rate is about 80%.
Figure 5: ROC curves generated from synthetic datasets with different temperature ranges and damage sizes. Plots (a-d) correspond to 2, 10, 30 and 60°C temperature range. Different curves in each plot correspond to different % damage reflections relative to the cut end reflection.

Figure 6: Area under ROC curves over different temperature ranges (on the horizontal axis) and different damage sizes (on the vertical axis): (a) mean (b) standard deviation

We can also use the ROC curves to determine the maximum tolerable EOC variations for a certain performance. For example, by comparing the grey solid lines representing 0.5% damage in plots (a-d), if we seek to have at least 80% POD and at most 30% PFA, the maximum temperature range in which we can still detect damage of about 0.5%, is 10°C. However, if we seek to detect the damage at 1% instead of 0.5%, we can tolerate a larger temperature range of 30°C.

Figure 6 clearly shows that as temperature range increases and damage decreases, the AUC decreases and its variance increases, except for in the upper right corner of Figure 6(b), where we seek to detect small damage under large EOC variations; this is because the baseline-subtracted residual is dominated by the noise.

In addition to being used as a way to estimate performance in different practical implementations, the ROC curves and the AUCs also provide a means compare different damage detection methodologies under varying test scenarios.
3. CONCLUSIONS

In this paper we demonstrated a cost-effective methodology to evaluate the performance of damage detection methods in various practical EOCs. By synthetically adding an artificial damage signal onto randomly selected undamaged experimental data collected in different EOC conditions, we can statistically evaluate our damage detection methods on different practical scenarios without costly creating real defects under all these EOCs.

We also demonstrated using ROC to illustrate the performance of damage detection. We showed the ROC of the temperature compensated, baseline subtracted damage residual in different test scenarios. We show that the ROCs can be used to estimate, for example, the POD and PFA under different EOC for specific damage size, or the maximum EOC we can tolerate to detect a specific damage size with specified POD and PFA.

In the near future, we plan to validate the synthetic superposition of damage using both finite element analysis and experimental data, especially for cases when defects are close to structural features that create complex reflected signals. We will also validate the ROC curves obtained from synthetic datasets with those obtained from experimental data, and apply the framework on other damage detection schemes such as SVD [5] and ICA [6]. These results will be presented at the conference and in subsequent publications.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the UK research centre in non-destructive evaluation (RCNDE) for funding this research through Engineering and Physical Sciences Research Council (EPSRC) grant (EP/L022125/1), and the Centre for Doctoral Training in Non-Destructive Evaluation Centre for funding an Engineering Doctorate studentship for Jacob Dobson (EP/I017704/1). The authors are also grateful to Guided Ultrasonics, Ltd. for providing the experimental data.

REFERENCES


