Author:
Karkera, Ben M

Title:
Structural health monitoring for marine applications using adhesively bonded piezoelectric transducers

Structural Health Monitoring for Marine Applications using Adhesively Bonded Piezoelectric Transducers

Benjamin Karkera

Department of Mechanical Engineering

UNIVERSITY OF BRISTOL

October 2021

A dissertation submitted to the University of Bristol in accordance with the requirements for award of the degree of Master of Science in the Faculty of Engineering

Word Count: 18,469

Abstract

The transition of Structural Health Monitoring to real-world marine applications remains rare, despite its significant potential for cost-saving. In the absence of established standards and best-practices, there is an increased emphasis on rigorous case-by-case qualification and validation of SHM systems, requiring an understanding of the individual components and methods, as well as the major assumptions and sources of uncertainty.

In order to transition to real-world applications, an important but challenging aspect to assess are the long-term factors affecting system performance. Representative trials are challenging and expensive, as the Structural Health Monitoring application systems they aim to emulate have operational lifespans of many decades.

An investigation is presented that evaluates a simple guided-wave ultrasonic-testing Structural Health Monitoring system. A hybrid Optimal Baseline Subtraction, Baseline Stretch Subtraction approach is adapted to a specific dataset to permit compensation for environmental and operating conditions. The results present two important design consideration for these systems: the exploration of noise suppression performance, the predictability of the behaviour of the system.

Additionally, the presence of a long-term change in the data was identified as the performance-limiting phenomenon observed in existing monitoring data that was collected over multiple years for the dataset used. A trial was undertaken to attempt to simulate long-term system aging effects in an accelerated manner, through the use of high temperature heat-treatment. The results indicate a promising relationship between the behaviours of the system and trends observed in data collected over many years.

In addition to technical considerations for application specific implementations and trials, discussion is given to the non-technical hurdles for industry implementation.

Acknowledgements

I would like to thank my supervisors, Professor Paul Wilcox and Professor Anthony Croxford. Also, Dr. Thomas Barber, Emily Connolly, and Wade Nixon for supporting me through this time.

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: BENJAMIN KARKERA   DATE: 09/10/2021

Table of Contents

Table of Figures

List of Tables

1. Introduction and Background

1.1. Problem statement

During the manufacture of naval vessels, Non-Destructive Evaluation – commonly referred to as NDE (or NDT for Non-Destructive Testing) – is carried out as a quality control measure.

NDE covers a broad range of methods, which make use of a variety of physical phenomena. Radiography (radiographic testing, or RT) measures the absorption of radiation passed through a component. Dye-penetrant inspection uses the different behaviours of a dye (caused by capillary action and adhesion and cohesion) in the absence and presence of dry powder to identify surface imperfections, such as cracks. Ultrasonic testing (UT) measures wave reflections and diffractions to detect material interfaces and discontinuities.

Raw material, castings, welded joints, and bends are subject to examination to ensure they are free from impermissible manufacturing defects. This examination does not change or damage the component. When a naval vessel is handed over by the manufacturer, an assurance is given that no defect is present above a critical size.

NDE at the build stage of the product lifecycle is focused on the detection of manufacturing defects. These defects, which result from failures or variations in manufacturing controls, include inclusions, laminations, and lack of weld fusion defects. In-service defects, on the other hand, which result from high stress, cyclic loading, or harsh environmental conditions under operation, typically exhibit slightly different characteristics, despite often initiating from manufacturing defects. These defects therefore may be suited to different NDE approaches. Corrosion, fatigue cracking, leaks and erosion are all in-service type defects.

Because of these differences (which may be subtle in many applications), manufacturing NDE and in-service NDE are often treated entirely independently. The asset is delivered with the assurance that no defect above a certain size exists, with the maintenance organisation then carrying out inspections to ensure no unacceptable defects have arisen due to operation. Only a small portion of the information generated during manufacture is passed on to aid in-service decision making, and research and development projects almost always focus on either in-service or manufacture independently.

It is the belief of many who work on the manufacture, operation, and maintenance of naval vessels, that a great opportunity exists for a more joined-up approach, where a holistic engineering view is given to the design and delivery of a serviceable product. Such an approach, and the techniques that enable it, are discussed in the following sections.

1.2. Permanently installed transducers

Permanently installed transducers may be directly bonded to target components at manufacture (or coupled using non-contact methods, for example as described in (Herdovics & Cegla, 2018)), for repeated inspection of area of interest without physical manipulation of the transducer.

Historically, the use of permanent installation of piezoelectric ultrasonic transducers has had many drawbacks. The requirement to expend transducers to monitor a relatively small area of the component, combined with an inability to scan, results in a high cost of coverage. Therefore, transducers tend to be more sparsely used, forcing each to cover a larger area, greatly reducing sensitivity. However, despite often causing a reduction in sensitivity and usefulness of each single measurement on its own, by permitting repeatable measurements to be made at regular intervals, permanently installed sensors have the potential to surpass point measurement (single point-in-time) NDE, greatly improving the ability to evaluate the remaining life of the component.

1.3. Continuous data collection

In addition to permanent installation of the transducer, another improvement over conventional point measurement NDE is the ability to regularly collect data in-service without a trained inspector being present or having to shut-down the system.

Recording whilst live, however, introduces a greater variability in environmental and operating conditions. While these do vary in conventional NDE, the low practical variation observed is usually within the bounds of error of the NDE technique used. In-service, high temperature systems may operate under wildly differing conditions, for example variation of many hundreds of degrees Celsius of working fluid within a pipe. The effects of these variations are often too complex for an inspector to interpret, and therefore new approaches are required to accurately evaluate the signals recorded.

The techniques used in this approach, and challenges presented, stray from those that are familiar in conventional NDE, and into the area of data science. Historic measurements can provide useful additional information that permits a better analysis of new data sets as they are recorded. This approach is known as Structural Health Monitoring, or SHM.

1.4. Structural health monitoring

There is a great deal of inconsistency in the use of terms for the technologies relating to monitoring, with Structural Health Monitoring, (Machine) Condition Monitoring, and Population (or fleet) Monitoring being assigned varying meaning, and sometimes being used interchangeably (Friedmann & Kraemer, 2016). For the purpose of this paper, SHM is treated as a subset of Condition Monitoring. That is, the continuous observing of a process, component, machine, or asset, for indicators of deviation from the nominal state (Condition Monitoring), with a specific focus on a load-bearing capacity of an item (SHM). SHM concerns typical load-bearing structure, such as a pressure hull, but also covers pipework, which is stressed by an internal and external pressure difference. Applications associated with the monitoring of rotating machinery (such as pumps, bearings, engines, or generators) are excluded from this work in order to limit the scope to purely SHM. However, there is significant scope for use of Condition Monitoring technology on a wide range of maritime applications (e.g. rotating machinery).

Often, the sensors used in SHM systems are similar to those used for conventional NDE, since the desired outcome of both is a strong correlation of structural damage to the physical phenomenon that the transducer is sensitive to. Differences in consideration between SHM and NDE sensor design exist, however, due to SHM transducers being permanently installed for local measurement. These include: temperature tolerance (SHM systems involve measurement while the asset is operational), data storage (long-term storage is needed), required operational life (the sensor should last the lifetime of the structure is it used to monitor), cost per transducer (as a transducer cannot be moved or reused), and housing (due to permanent bonding to the structure).

Operational monitoring systems may also be employed, where sensors infer damage indirectly through the correlation of other measurements (e.g. loads on the structure, or environmental conditions) (SAE International, 2013). These systems are not considered for this work.

SHM is heavily dependent on individual system specific data. This increases the reliance on historic data or a start-of-life baselining period, to permit background subtraction, or independent secondary measurements for multiparameter analysis.

The environmental and operating conditions experienced by an asset and monitoring system will usually lead to variations in the sensor response. This can often be large and may mask a change in signal due to actual defect growth. Being able, then, to compensate for these changing states allows for improvement of defect detection (Croxford, et al., 2007). One method of achieving this is by collecting a full library of defect-free signals from the signal

in every operational and environmental condition. This means the system contains a model of the "normal" or undamaged state, permitting it to then flag up any abnormal state, which can indicate that the structure is potentially damaged.

The primary ingredients for monitoring, for structural health or otherwise, are: sensors, measurement instruments, signal processing methods and thresholds for alarms (Friedmann & Kraemer, 2016). Each of these components may require different levels of development work to meet the requirements of the monitoring system. In many cases, these components may be investigated independently, however, final system-level validation is crucial to evaluate the complex interactions of each component and the unique effects that will be introduced by each application.

Due to the challenges, complexities, and application-specific nature of SHM systems, as much effort should be given to verification and validation as to the design. Sensitivity, sources (and magnitude) of uncertainty, validity of assumptions, and long-term effects must be considered if these systems are ever to be use in lieu of current NDE.

Validation of these techniques, where data-based "black-box" methods are often employed, often requires an approach different to that commonly used in NDE qualification. Because behaviours of the system are not based as strongly on physical reasoning, there is a higher dependence on statistical performance measures.

1.5. Gaps in literature

The most significant gap in the development of this technology is transition to real-world applications. Despite the surge in the volume of research in Structural Health Monitoring over the past decade, there are only a few implementations of the technology within industry (Cawley, 2018). There is great value in work evaluating the higher technology maturity aspects, such as validation, component evaluation and discussion on practical considerations for application-specific design.

Although in recent years a substantial number of standards have emerged for the use of Condition Monitoring of rotating machines, there are still very few widely accepted standards, industry practices or validation methods for Structural Health Monitoring. A notable exception to this being a standard written by SAE International for fixed-wing aircraft (SAE International, 2013).

In the absence of standards, there is an increased emphasis on rigorous case-by-case qualification and validation of SHM systems, requiring an understanding of the individual components and methods, as well as the major assumptions and sources of uncertainty.

In order to transition to real-world applications, an important but challenging aspect to assess is the long-term effects of environmental and operational factors that lead to changes in system behaviour over significant periods of operation. SHM systems need to operate consistently for many years, which makes representative trials less practical. Studies have been performed to understand the long-term stability of guided wave sensors (Attarian, et al., 2014), and shown the importance of changes in the adhesive bond. Therefore, the behaviour of the adhesive (and the specific process by which the adhesive is applied) should be understood as a source of change in ultrasonic response.

1.6. Project overview

1.6.1. Aims

The aim of this work is to improve the understanding of certain aspects of a specific guided wave structural monitoring system. Although the focus will be on a specific dataset and system, it is believed that the methodologies and discussions will be relevant for the extension to other applications and datasets.

1.6.2. Investigation of training and baseline subtraction performance on long-term data

The first part of the analysis aspect of the project used historic data, for the development of a generic data analysis system, as well as provide the basis for how performance would be measured. It would not have been feasible to have collected this data in the timescales of this project, therefore, the only practical approach was to decouple the data acquisition (hardware) and analysis (software) components of the system.

The historic data used provided a sufficiently long timeframe required to provide effects expected in a real industrial asset and monitoring system's life, including the impact of age dependent effects.

The developed monitoring system was evaluated for performance and validity with practical monitoring considerations. The dependence of the operations used to compensate for EOCs were investigated to identify computational savings that could be made. These reductions in computational burden make data-heavy approaches more practical on hardware that can be embedded onto assets. Appropriate ranges for each operation were determined through subtraction performance, and consideration of realistic EOC ranges to be compensated.

Finally, a discussion of overall algorithm subtraction performance and computation time using different ranges, implementation of operations, and computation savings was used to highlight the need for application-specific tailoring.

1.6.3. Practical assessment of temperature cycling on sensors and bond

A simple guided wave hardware system was constructed on a number of small metal plate samples. On these samples, a trial was undertaken to assess the contribution of sensor system aging representative of that experienced by the equipment and bonding used to collect long-term data. The trial, however, aimed to replicate this in an accelerated manner through the use of high temperature cycles. The aims of this trial were to measure bandwidth, centre frequency and amplitude of a signal collected on each of these samples as representative measures of performance, whilst attempting to induce an accelerated aging effect through high-temperature cycling.

The trends in each of these variables was assessed during regular recordings over a heat-treatment period of days, as well as a comparison to single final measurements made after a much longer heat treatment. Additionally, the heat-treatment effect was assessed using the signal difference metric, detailed in the Training and Baseline Subtraction section.

2. Selection of Measurement Technology

2.1. Monitoring technologies

Due to SHM being concerned with a component's structural integrity, the sensors and measurement systems are often similar to those used in single-point-in-time NDE. Most structural health monitoring sensors use active excitation, with the exception of acoustic emission and thermal/visual imaging techniques. Passive monitoring tends only to be appropriate in applications where detection of only very severe damage is necessary, due to low sensitivity that is typically achievable.

For local monitoring, where the intention is to monitor known damage or sites of expected failure, it may be economical to monitor a small volume near to the transducer (Cawley, 2018). Bulk wave Ultrasonic Testing may be the most appropriate solution, providing high sensitivity, and simplicity.

Ultrasonic guided wave testing is a promising candidate for large area, in-service monitoring (with a limited number of sensors) for structural assessment. For this reason, guided wave UT was selected for this work. Additionally, this is an active research area, with promising signs of adoption and development for industry. It should be noted that in certain applications, where damage is expected to be uniform, local monitoring may be used to cover large areas by assuming equivalence (Cawley, 2018).

The main target defect types for marine applications are corrosion (both uniform and local) and through-wall cracking, for which UT is recognised as one of the most reliable NDE methods for detection.

2.2. Choice of method: Guided wave UT

Ultrasonic testing (UT) is the practise of passing high frequency (typically 0.1 – 10 MHz) mechanical waves through a material, which reflect (and diffract) off interfaces, allowing the identification and location of these interfaces, which may be defects or component geometry. Guided wave testing is a branch of UT that refers to the use of wave modes that are constrained by the surfaces of components. Differing from bulk waves, less energy is lost to beam spread, as the material interfaces maintain the direction and mode of the wave in fewer dimensions, without any meaningful transmission of energy through the free surface. These waves, therefore, can travel long distances in plates or pipes, until their path is interrupted by a change in the component. Because guided ultrasonic waves travel parallel to the plate surface, they interact most strongly with defects oriented perpendicular to the plate surface, such as through-wall cracks or local wall thinning (due to corrosion).

Certain conditions are required to generate, and effectively use, these guided waves. Therefore, inspection frequency selection must be considered more carefully for the specific component wall thickness than for bulk-wave UT.

Guided wave UT has been an established inspection tool in other industries since the 1990's, with its main applications being large-area inspection of buried pipelines for the Oil & Gas industry (ASM Handbook). The ability to inspect many tens of metres of pipework (ASM Handbook) from a single point, greatly reducing time and cost of inspecting inaccessible assets. A significant amount of success has been demonstrated on pipes and rails (one-dimensional structures), where component complexity is low, and walls may be used as a wave guide (Cawley, 2003). Commercial systems have been demonstrated to be capable of detecting changes in cross section of around 5% (Cawley, et al., 2012). However, with insufficient capability to accurately evaluate loss of material, the inspection is often complimented by higher-accuracy thickness gauging or NDE after guided wave screening has identified thinning.

The greatest limitation of guided wave inspection is the complexity of measurements, with large coverage resulting in many signals merging, obscuring any contribution from damage. This challenge for signal interpretation means that guided-wave NDE is only typically suitable for simple structures with few features (e.g. pipes and rails). In these guided-wave systems, a significant amount of effort must also be put into the design of transducers and signal processing so that no unwanted wave modes are excited, and that any excited modes are non-dispersive (Alleyne & Cawley, 1992).

The benefit to naval vessels of guided wave inspection is in principle good, with large areas able to be screened for damage with lesser access requirements. However, because of the high feature density of naval vessels and structure, GWUT inspection is generally not practical.

3. Development of a Training Algorithm for GWUT Data

3.1. Overview

This section details the approach taken for the design of a training algorithm used to process GWUT data, detailing the development of an approach specific to an example dataset, also considering extension and generalisation to other applications.

First, details of the input data are given, as an understanding of this influences the design of any processing step. Next, an outline is provided for the overall method used for training an algorithm to automatically flag indicators of damage. Subsequent sections evaluate the specific signal processing and decision-making components in turn, and provide details of the design of the final algorithm developed.

Signal processing is the one of the most challenging aspect of the design of a monitoring system and is where a large portion of the system design effort should be concentrated. This is where the most difficulty exists in transitioning to highly variable industrial environments and applications, with the ability to suppress benign-feature signals being a prime determiner of the performance of guided-wave inspection (Croxford, et al., 2007).

3.2. Dataset investigation

The first step in developing a training algorithm for monitoring is to understand the input data.

A significant challenge with SHM work is the availability of representative data for real structures under realistic environmental conditions collected over timescales that reflect those expected in monitoring applications.

A dataset was available for use in the project, which had been collected from a structure at the University of Bristol (Figure 1) between 2012 – 2015. This provided a dataset that would not have been practical to collect as part of this project.

An array of transducers was bonded onto the structure, a ~2 m diameter steel tank, located outdoors, with a data acquisition instrument located nearby, collecting data at various times of day. Despite data being available for the full array, for this work, only single pairs of transducers were used.
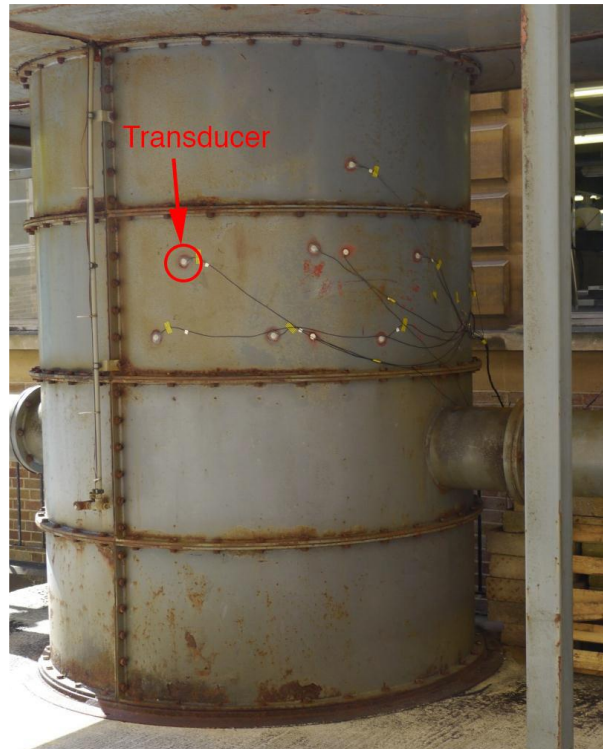
Figure 1 Transducers bonded to a tank outside the University of Bristol use for data collection over a number of years (Courtier, 2018)

As this data was recorded on a structure primarily affected by environmental conditions, the rate of change introduced is expected to be low, with the highest frequency change occurring due to the day-night cycle. Recording times were varied to capture a range of times-of-day over a 4-year period. Figure 2 shows the distribution of collection dates and times of the data.
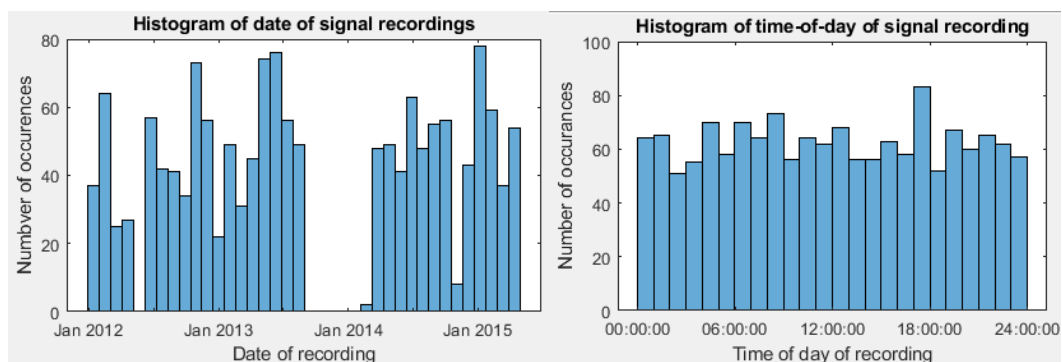
Figure 2 Histograms of 1500 signal recording dates (left) and times (right), with bin widths of one month and one hour, respectively.

As shown, the data was collected over a three-year period, with coverage of different times-of-year, giving both a significant time-period to give indicate age dependent effects, and represent a wide range of environmental conditions. The break in readings is not expected to have a significant effect on the usefulness of the data for this work.

3.3. Training algorithm

3.3.1. Overview

To be able to detect and flag abnormal asset states, it is necessary to build an efficient, but complete, model of the undamaged state of the asset. For this system, the model is described by a library of baseline signals that represent any measurement made while in an undamaged state under any operating condition. To determine the level of similarity between new signals recorded during operation and the baselines, signal subtraction is used.

Temperature is understood to be the dominating factor affecting guided wave signals. Although theoretically it is possible to use direct temperature measurements to determine and remove temperature effects on a signal, it is generally not practical in real applications. Inaccuracies in temperature compensations of only a few degrees can mask target damage signals (Lu & Michaels, 2005). Accurate temperature measurement of even very localised areas is notoriously challenging in practical applications. This application introduces a further complication that would be impractical for temperature measurement: instances where temperature is non-uniform along the component, and therefore, guided wave propagation path. Additionally, these temperature compensation techniques only remove the primary effect of temperature, and do not account for the additional effects on the system. For example, (Putkis, 2014) observed significant changes in transducer and bonding behaviour with temperature, compared to the effect of temperature on sound propagation in the material itself.

Practical temperature compensation techniques, therefore, operate by focusing on the problem of effective removal of the effects of unknown and unmeasured temperature variations (as well as other EOCs). It should be noted that modelling temperature conditions is more complex than considering the state at a range of global temperatures. Local temperature differences and hysteresis result in complex states and interactions that make predicting temperature effects on a structure challenging, and limited in how much can be transferred between different systems with different temperature conditions and responses.

3.3.2. Optimal Baseline Subtraction

The simplest method for achieving this is the Optimal Baseline Subtraction (OBS) method, as detailed in (Croxford, et al., 2010). This involves a training phase, in which the system is assumed to be damage-free, and recordings taken until there is sufficient confidence that all possible conditions have been observed. Each of these recorded signals are compared to the baseline set (the set of signals that are used to represent the damage-free state of the system). If a new signal is determined to be different from any signal in the baseline set, and the system is still assumed to be in a damage-free state, then a new state has been found. This

new signal is added to the baseline set, with the model of undamaged state becoming more complete. The technique, if successful, should add fewer baseline to the reference set as time goes on, as it should become less likely that new conditions are encountered. Strictly, after the system is cycled through its entire range of operating and environmental conditions, the baseline reference set should cease to grow. This does assume, however, that there is no ageing, or progressive change, of the system over time. This aging is commonly observed on the transducer and bonding. If this does occur, there is a secondary change to the signals occurring at the same time as the potential damage growth. For this method to be useful, there must be confidence that the measurement system will remain largely unchanged over the system life.

The training phase is also critical to the performance of the system. The assumption that no permanent change to the structure occurs during the training phase is a significant one. As well as the limitation that, to be accurate, the system must be subjected to the entire (or even a slightly larger) range of operation conditions to those feasibly experienced during operation. This may not be practical in some cases, and so other methods may be preferable, which permit a wider range of conditions to be generated from a smaller set of training data.

3.3.3. Baseline Stretch Subtraction

A second method, Baseline Stretch Subtraction (or BSS), uses only one reference baseline. The principle of this technique is that: by performing mathematical operations on this reference signal, it should be possible to mimic the response in the system at a broad range of environmental and operating condition (within certain bounds). A description of this method can be found in (Croxford, et al., 2010).

The baseline stretch process requires the performing of mathematical operations on the ultrasonic signal. These operations model the effects of environmental and operating conditions on guided wave signals. Performing these operations on analytic signals (example given in Figure 3) in the frequency domain often simplifies these operations.
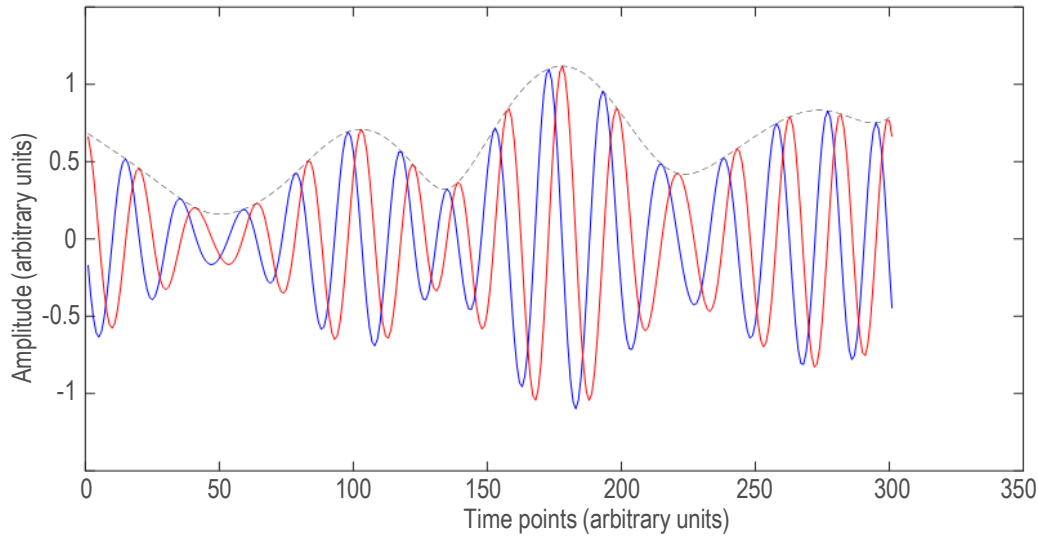
Figure 3 Real and imaginary components of complex analytic signal for an arbitrary section of recorded signal (blue and red solid lines respecitvely); and Hilbert envelope (dotted line).

The following operations are used to compensate for the predicted phenomenon that occur during changing operating conditions in the "non-damaged" state. Therefore, baselines covering an entire range of operating conditions may be synthesised from a single measurement.

- A scaling of amplitude of the overall signal is used to model the temperature-dependent effect of changes in coupling/bond.
- Time domain translation may be used to align signals to accommodate for any jitter. "Jitter" is the term used for when the data acquisition instrument causes a discrepancy between different measurements in where the emitted pulse occurs around time = 0 on the time trace.
- A phase shift operation may be used where it is suspected that transducer behaviour may change with age. The objective is to shift the phase of the signal without altering the Hilbert envelope.
- A time stretch operation to model temperature effects on wave-packet arrival time.

The primary effect of system temperature on a pitch-catch measurement between the two transducers, is to change the velocity of the wave-packets. This has the effect of increasing or decreasing the time taken for a wave-packet to reach the receiver. Although there is a secondary effect that the thermal expansion of the component changes the distance between the two transducers, it has been shown that this effect is much smaller than the velocity change, and so can be ignored (Croxford, et al., 2007). An approximation of this effect is the stretching of the time axis of the signal, to change time of arrival of wave packets. The true effect of a change in temperature is a translation of the received signal in time, rather than

13

stretching or compression of the entire signal in time (Croxford, et al., 2007), and so this operation is only valid for small temperature variations. Additionally, with stretching and compression, the frequency content of the wave packet will be altered. A typical appropriate range of compensation of $1 - 5°C$ is given in (Clarke, et al., 2010).

The fact that these operations are only valid for compensation of small variations exposes the drawback of the BSS technique. Conveniently however, the two techniques (OBS and BSS) may be used together to provide a hybrid method that combines the advantages of both.

3.3.4. Hybrid approach

A hybrid method combining the OBS and BSS techniques may be used where multiple baselines are collected, and also stretched/shifted to fill any gaps in the collected baselines set.

Figure 4 is a diagrammatic representation of the initial simple implementation of the hybrid OBS + BSS approach, where each new signal is transformed using every combination of the four stretch parameters before each of these transformed signals is compared to a set of baseline measurements to find the closest overall match. Whether the difference metric between the closest matching transformed signal and baseline is below a defined threshold determines whether the signal is flagged as "normal", or as potentially containing damage. Given the large number of combinations of these four stretch parameters, computational savings and efficiency improvements are especially important.

A provisional measure of signal difference was implemented: the maximum amplitude of the residual signal remaining after the signals were subtracted from one another. To achieve this, the signals were transformed using all combinations of stretch parameters, and the best match (minimum difference metric) found.

This could be considered a brute-force approach, and since behaviour is not controlled, there is less inherent confidence in the system behaving as expected under all input conditions. Therefore, there is an increased onus on verification and validation of results and behaviours. It is important to ensure that only plausible values of each parameter are used, such that all outcomes may be explained by some real-world phenomenon. The simplest example of this being, that the responses of the system are comparable for signals recorded at similar operating conditions.
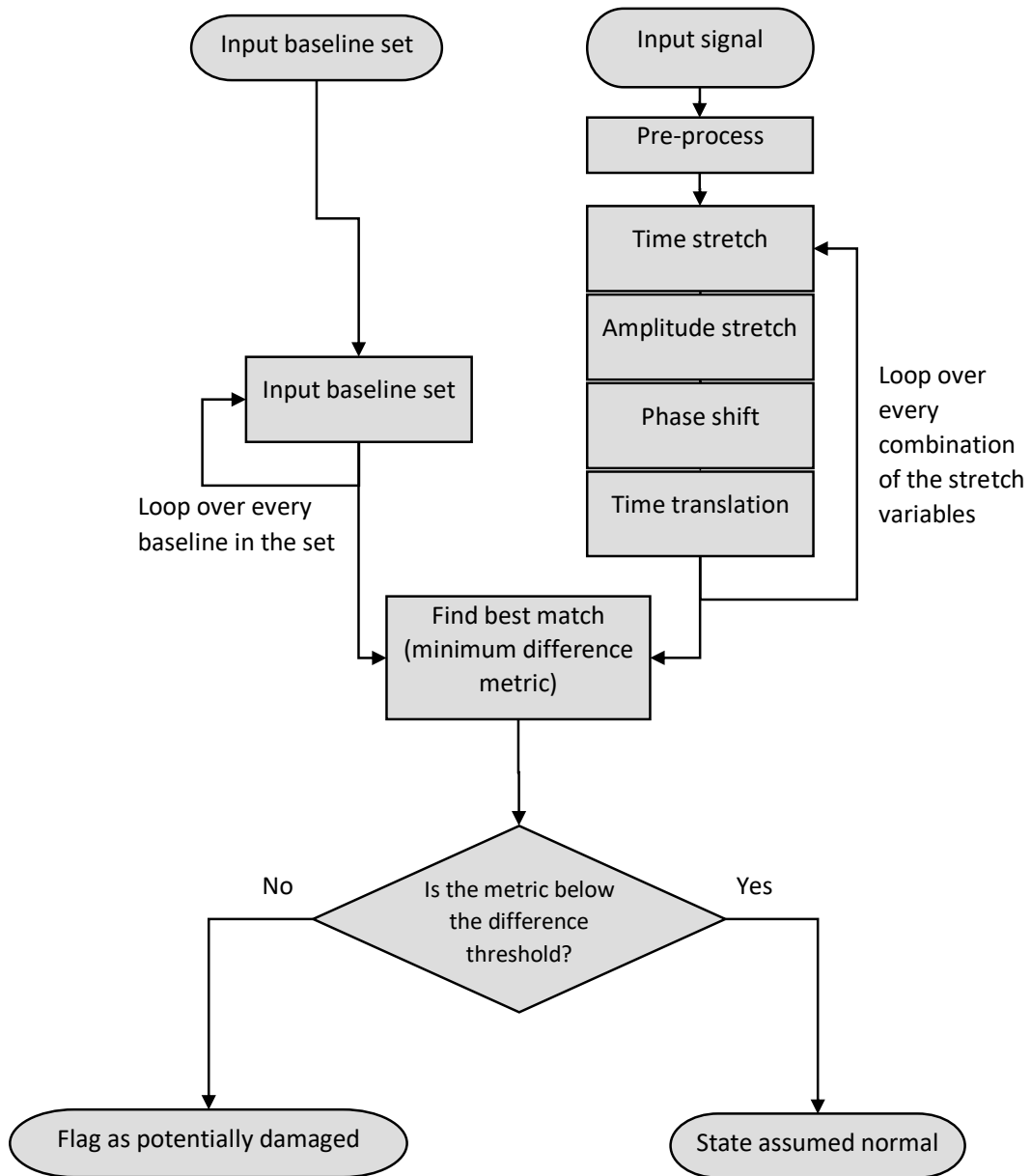
Figure 4 Flow diagram describing hybrid approach to training.

3.4. Baseline set reduction

The objective of this approach is to provide a set of signals that can be shifted to cover any undamaged state of the structure with any combination of EOCs. There will undoubtedly be baseline signals recorded, however, that are redundant – covering approximately the same EOC state of the structure.

Due to the computational burden of calculating all combinations of baselines under all combinations of shift states, it is beneficial to reduce this baseline set to a more efficient representation of the undamaged state. It is possible to write an algorithm capable of comparing each of the baselines in the originally recorded set to each other – in a similar

way to the approach described in Figure 4, however, with dissimilar signals being added to a final/reduced baseline set (instead of the output being "Flag as potentially damaged" in Figure 4), and similar signals being discarded due to being redundant (instead of the output being "State assumed normal" in Figure 4).

An additional purpose to development of this algorithm is the ability to gain insights into the behaviour of the system at different EOCs. By comparing the subtraction performance when using different amounts of shifting, and using different algorithm implementations, it was possible to identify improvements to computation time, as well as understand system behaviour, which are important factors for SHM system qualification and implementation.

3.5. Pre-processing

Now that the outline of a hybrid OBS+BSS approach has been discussed, the subsequent sections detail development work to optimise the algorithm for the data available.

The first of these steps is the sanitising of input data, and normalisation, before the main processing algorithm is carried out. Pre-processing is a vital step in any data processing method for improving performance through removal of unwanted information and ensuring suitable formatting of data.

3.5.1. Normalisation

Normalisation is used to both simplify processing operations, and to generalise the code for different inputs to make it easier to adapt to different applications.

Amplitude normalisation is performed by setting a known signal feature to unity. In this application, it was chosen to set the first arrival signal amplitude to a value of one. As the first arrival signal is also used as the reference amplitude, this further simplifies subsequent signal analysis.

A part of time base normalisation is performed by cropping data at pulse emission time (time = 0). The purpose of this is to simplify time-domain operations. For example, time domain stretching can be performed about the time where distanced travelled is zero, by stretching about data point one. No information is lost in cropping the data at this point, as any signal before time = 0 (pulse emission) can be assumed to be noise, with the first arrival of importance arriving at a time corresponding to: $time = sound\ velocity \times transducer\ separation\ distance$. It should be noted that in practise, instruments do not perfectly map transmission time to time= 0, and therefore some compensation needs to be made. This is handled later in the processing procedure.

Time-base normalisation can be used to generalise the code when sampling rate and gate length vary across applications. As only one dataset was used in this work, time-base normalisation was not implemented.

3.5.2. Digital filtering

A significant part of pre-processing is the removal of unwanted or unneeded information, without loss of useful information. Digital frequency filtering is used to selectively discard information based on signal frequency. The centre frequency and bandwidth of the input signal is (approximately) known, and it can be assumed that any contribution outside this frequency range is unwanted noise. A Gaussian bandpass filter was used to smoothly remove high and low frequency noise. The centre filter frequency used was the nominal input frequency of 250 kHz, with a filter Bandwidth of 100 kHz (as shown in Figure 5).
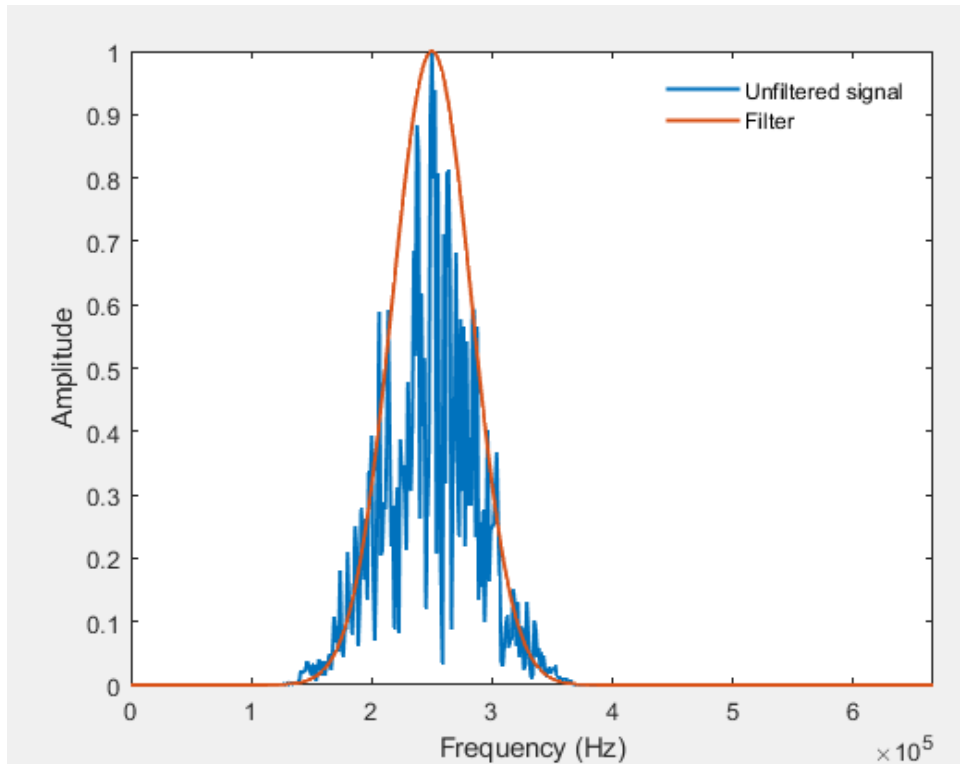
Figure 5 Normalised amplitude vs frequency plot of signal frequency content before filtering, with filter overlayed

It is also beneficial to remove any time points that are known to be of no use to the inspection, due to being outside the feasible range of detectible defect. Any time point before the first arrival signal can be disregarded. Additionally, the time, or more specifically distance travelled, after which the ultrasonic response is no longer of value can be determined through consideration of required coverage, attenuation (both material and beam spread, etc.), and noise. A Tukey Window was used to smoothly crop the data between two time points (Figure 6). This cropping method provided a smooth transition of amplitude to

zero at either end of the data, enabling correct implementation of the Fast Fourier Transform operation. The transition time of the Tukey window is determined ensuring five periods of the centre frequency have passed between the rise start and end (and fall start and end). For a typical centre frequency used in this work of 250 kHz, the transition time T is:

$$T = 5 \times \frac{1}{250\ kHz} = 20\ \mu s \tag{1}$$
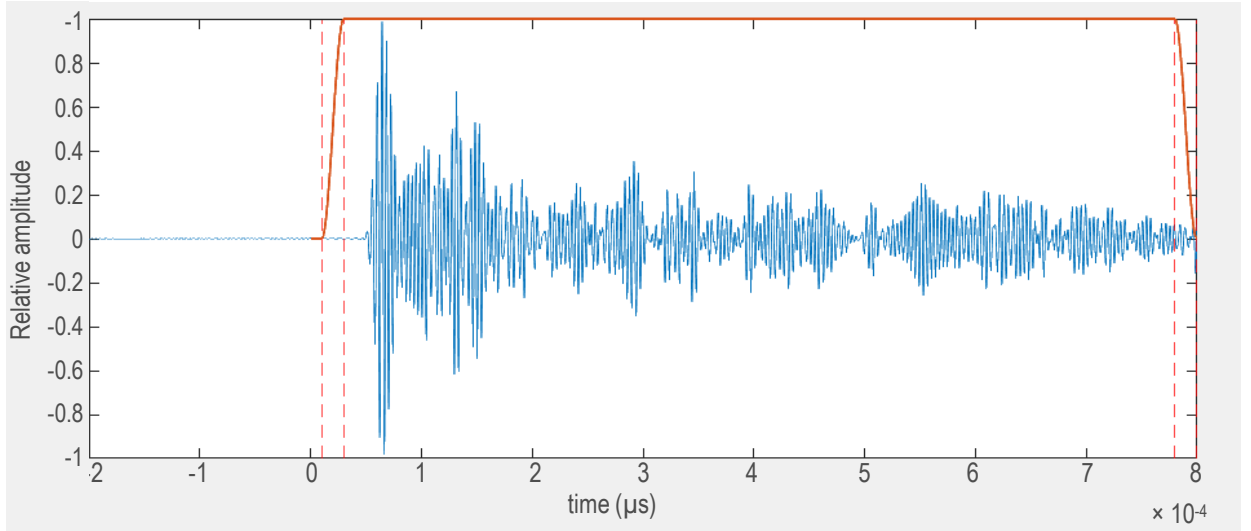
Figure 6 Tukey window (orange) used for time-domain cropping of recorded signal (blue) using "rise" and "fall" starts and ends shown by dotted vertical lines.

3.5.3. DAC

Distance amplitude correction (DAC) is carried out to accommodate for losses in amplitude due to beam spread over the propagation distance of a wave packet. By correction for this loss of signal amplitude, the hope is that a similar reflector at any path length will give a similar response, meaning sensitivity is even over the entire monitored range.

A consequence of this distance amplitude correction is that any noise is also amplified, meaning signals at longer time paths will have an increased noise amplitude (i.e., signal to noise ratio at any point in the recorded time base is unchanged by DAC).

Care should be taken with applying DAC, as the assumptions made in their calculation will only be appropriate for certain regions of the signal time domain. The case described here is a system whereby a single pair of transducers covers a large detection area, multiple times the transducer separation distance. A long time-range is recorded, giving sensitivity to far-away regions of the component, where beam losses are considerable. DAC is therefore sensible. For a component with a denser covering of transducers (each covering only up to

two times the transducer separation) DAC may not be required, and if used, may introduce significant error.

In the material and configuration used, losses due to beam spread are expected to be the main source of amplitude loss with propagation distance. The other main source of energy loss, material attenuation, would be a significant factor in other structural materials. It is proposed that an additional amplitude loss factor:

$$Amplitude \ loss \ factor = e^{-\alpha d} \tag{2}$$

(where α is the attenuation coefficient, and d is the propagation distance), could be used to accommodate for this loss (Krautkrämer & Krautkrämer, 1990). The presence of significant material attenuation would result in a further decrease in useful detection range.

Due to the use of a small transmitting transducer, the losses due to beam spread are approximated using the inverse square law model. In a two-dimensional plane, (in this case a thin plate,) this states that for an initial energy per unit length (E_1), the measured energy per unit length (E_2), at any distance away (d), is given by:

$$E_2 = \frac{E_1}{d} \tag{3}$$

Since energy per unit length is proportional to the square of the signal amplitude (A) at a receiver, this can be rewritten as:

$$A_2^2 = \frac{A_1^2}{d}, \qquad A_2 = \frac{A_1}{\sqrt{d}} \tag{4}$$

The only spatial information contained in single-pair transducer signals is a time of flight for each wave packet. Assuming a single speed, this corresponds to a physical total propagation distance of the wave packet. For a hypothetical pair of transmitter and receiver transducers (Tx and Rx, respectively, in Figure 7) and a single reflector, there is an ellipse of equal propagation distance that can be determined. A reflector at any point on this ellipse would give the same time of flight on the recorded A-scan. Although for a given time of flight, the *total* path length is equal, there is a difference in the ratio of each of the *individual* paths (see Figure 7). This difference in paths results in differing loss of energy due to beam spread.

In this instance, it is impossible to determine the true ratio of individual paths (d_{1-2} and d_{2-3} in Figure 7), or therefore, the true loss due to beam spread. A single approximation must therefore be selected, which includes an inherent error.

This section details the determination of an order of magnitude error from the assumed DAC model used.
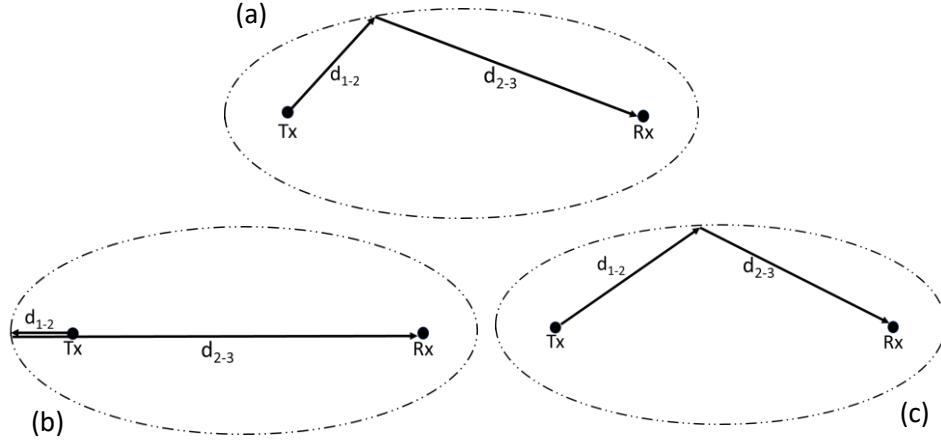
Figure 7 (a) Hypothetical elipse (dotted line) for single reflector lcoation that would give an equal path distance (solid line) between transmitter and reciever. (b) Path that gives minimum beam spread for a given total path length. (c) Path that gives maximum beam spread for a given total path length.

Assuming the defect (D) act as a point reflector, for a total path length of $d_{1-2} + d_{2-3}$, the total loss due to beam spread between transmitter T_x and receiver R_x is given by:

$$A_D = \frac{A_{T_x}}{\sqrt{d_{1-2}}}, \qquad A_{R_x} = \frac{A_D}{\sqrt{d_{2-3}}} \tag{5}$$

$$A_{R_x} = \frac{A_{T_x}}{\sqrt{d_{1-2}d_{2-3}}} \tag{6}$$

where the term A_D is the amplitude at the defect.

The DAC compensation factor (α) necessary can be calculated by the amplitude at transmission divided by the amplitude of the signal at the receiver.

$$\alpha = \frac{A_{T_x}}{A_{R_x}} \tag{7}$$

$$\alpha = \sqrt{d_{1-2}d_{2-3}} \tag{8}$$

Inherent in this assumption is the lack of reflections from benign structural features that will cause additional diffraction and beam spread.

The maximum beam spread occurs when the two paths are equal ($d_{1-2} = d_{2-3}$), as shown in Figure 7 (c).

$$\alpha_{max} = \sqrt{d_{1-2}d_{1-2}} \tag{9}$$

$$\alpha_{max} = d_{1-2} \tag{10}$$

Or, in terms of total path length d:

$$\alpha_{max} = \frac{d}{2} \tag{11}$$

For a given total path length, the minimum beam spread occurs when either individual path (d_{1-2} or d_{2-3}) is maximised, and the other minimised (as shown in Figure 7 (b)). For example, when d_{1-2} is minimised:

$$d_{2-3} = d_{T_xR_x} + d_{1-2} \tag{12}$$

where $d_{T_xR_x}$ is the physical separation of the transducer pair, and is therefore a constant.

$$\alpha_{min} = \sqrt{d_{1-2}\left(d_{T_xR_x} + d_{1-2}\right)} \tag{13}$$

To give a meaningful result in terms of measurable variables, the total path distance d can be substituted:

$$d = d_{T_xR_x} + 2d_{1-2} \tag{14}$$

$$d_{1-2} = \frac{d - d_{T_xR_x}}{2} \tag{15}$$

Therefore,

$$\alpha_{min} = \sqrt{\frac{d - d_{T_xR_x}}{2}\left(d_{T_xR_x} + \frac{d - d_{T_xR_x}}{2}\right)} \tag{16}$$

$$\alpha_{min} = \sqrt{\frac{d - d_{T_xR_x}}{2}\left(\frac{d + d_{T_xR_x}}{2}\right)} \tag{17}$$

$$\alpha_{min} = \frac{1}{2}\sqrt{d^2 - d_{T_xR_x}^2} \tag{18}$$

Considering a single velocity (v), the DAC compensation factors (α) can be given in terms of time of flight (t_{of}).

$$\alpha_{max} = \frac{vt_{of}}{2} \tag{19}$$

$$\alpha_{min} = \frac{1}{2}\sqrt{(vt_{of})^2 - d_{T_xR_x}^2} \tag{20}$$

Using the Maximum as a conservative approximation, the error (ϵ) of DAC compensation factors (α) is between:

$$0 > \epsilon_\alpha > \alpha_{min} - \alpha_{max} \tag{21}$$

Alternatively, the true DAC value (α), written:

$$\alpha = \alpha_{max}\begin{pmatrix} +0 \\ -(\alpha_{max} - \alpha_{min}) \end{pmatrix} \tag{22}$$

A limitation of this model is the simplification of reflections as point reflectors, therefore not accounting for real differences in reflector morphology and orientation.

As can be seen from Figure 8, this error tends to zero (and where Maximum and Minimum approximations converge) when the square of the path distance $d^2 \gg -d^2_{T_xR_x}$, and so the ellipse (Figure 7) approaches a circle.

It should be noted that the plot starts at path distance equal to 1, and not zero, as no valid signal is expected to arrive before this time in pitch-catch modality.
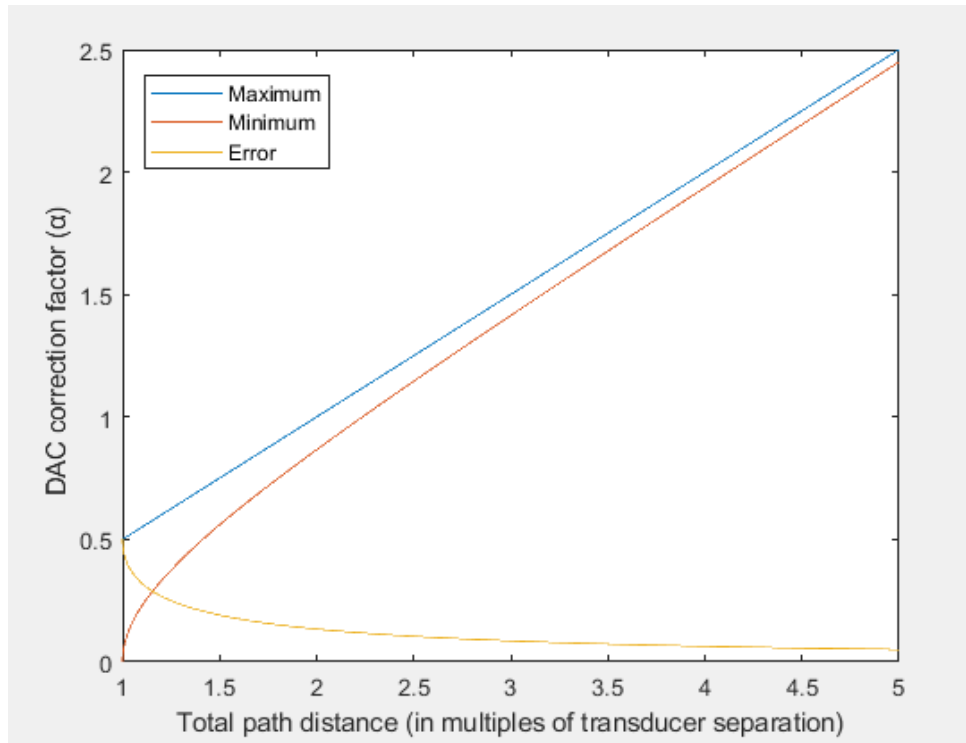
Figure 8 Maximum (blue) and minimum (red) DAC values as a function of total path distance, given as multiples of the transducer separation.

The possible error is substantial near to the direct arrival path (path distance $d = 1$). It is worth discussing this area of the plot, to evaluate whether this is significant to the validity of

the technique over its assumed coverage range. First, the validity of the Maximum model is considered.

This DAC calculation is based on the compensation of two paths, each contributing an inverse-square law beam spread factor. It should be stated, then, that this is not appropriate for the first-arrival signal, for which the beam-spread is considerably lower. For the direct path of the first-arrival signal, the DAC is given by the model:

$$\alpha_{direct} = \sqrt{d_{T_x R_x}} \tag{23}$$

Given that it is not practical to compensate the first arrival signal differently to other signal contributions that have a similar time of arrival, it may be beneficial to consider the case where both models match (at distance $d = d_{T_x R_x}$):

$$\sqrt{d_{T_x R_x}} = d_{T_x R_x} \tag{24}$$

This occurs where the transducer separation $d_{T_x R_x} = 1$.

Since physical separation of transducers will be set by practical design considerations, the only practical method for achieving this, is to use the separation distance as the measurement unit.

The other potential ways of removing this issue of incorrect first-arrival amplitude correction would be cropping out, or differently scaling, the first arrival signal contribution in the residual signal. This would, however, result in incorrect scaling of signals in this region, or the complete loss of valuable information about the first arrival amplitude, for example a change due to defects blocking or altering the first arrival signal.

In summary, by using a measurement unit of the transducer separation, error of first arrival distance amplitude correction can be effectively removed. For this reason, DAC values in Figure 8 are given as functions of path length in multiples of transducer separation. Therefore, the Maximum model in Figure 8 is most accurate at path length $d = 1$.

If the Maximum model is accurate in this region, and there is large difference between the minimum and the maximum near to path length $d = 1$ in Figure 8, discussion of the Minimum approximation is necessary. The tending of the Minimum approximation to zero is due to the collapsing to zero of one of the distance terms in $\alpha = \sqrt{d_{1-2} d_{2-3}}$, which is not encountered when considering the Maximum beam spread approximation, where neither path lengths can be less than half the transducer separation, and can therefore never approach zero.

The minimised distance $d_{minimised}$ is the problematic term that leads to invalid compensation factors. The amplitude correction value with two reflectors should never be larger than that for the first arrival signal.

For the Minimum model, where:

$$\alpha_{min} = \sqrt{d_{1-2}\left(d_{T_xR_x} + d_{1-2}\right)} \tag{25}$$

To find the distance d_{1-2} for which the Minimum model is valid, the compensation value must be greater or equal to that of the first arrival correction.

$$\alpha_{min} \geq \alpha_{direct} \tag{26}$$

$$\sqrt{d_{1-2}\left(d_{T_xR_x} + d_{1-2}\right)} \geq \sqrt{d_{T_xR_x}} \tag{27}$$

$$d_{1-2}^2 + d_{T_xR_x}d_{1-2} - d_{T_xR_x} \geq 0 \tag{28}$$

$$\left(d_{1-2} + \frac{d_{T_xR_x}}{2}\right)^2 - \frac{5d_{T_xR_x}}{4} \geq 0 \tag{29}$$

$$\left(d_{1-2} + \frac{d_{T_xR_x}}{2}\right)^2 \geq \frac{5d_{T_xR_x}}{4} \tag{30}$$

Taking the positive solution:

$$d_{1-2} \geq \frac{\left(\sqrt{5} - 1\right)d_{T_xR_x}}{2} \tag{31}$$

The validity of the Minimum correction model improves with the use of smaller units of distance measurement. However, taking the unit selected to make the Maximum correction model valid for the first arrival signal (where $d_{T_xR_x} = 1$), for the distance:

$$total\ path = d_{1-2} \times 2 + d_{T_xR_x} \tag{32}$$

The Minimum correction model is valid at a total path:

$$total\ path \geq \left(\sqrt{5} - 1\right)d_{T_xR_x} + d_{T_xR_x} \tag{33}$$

$$total\ path \geq \sqrt{5}d_{T_xR_x} \approx 2.2\ d_{T_xR_x} \tag{34}$$

The region on Figure 8, where the total path is close to the transducer separation is where the greatest potential error exists. Using the Maximum model will give a greater correction factor than the true correction value.

There are certain factors that mean the practical effect of this is lessened. The presence of the first arrival signal itself, which is by far the largest expected signal, will mask indications and introduce a high degree of variability in this area. It may be beneficial, then, for indications close to the first arrival to have a potentially conservative (larger than true ideal) distance amplitude correction factor.

The range of inspection is limited by the DAC noise floor. The first arrival signal occurs at distance = 1 (when in units of transducer separation). This is used as a reference, and occurs at a predictable distance, so any measurement can be easily compared to first arrival.

By recording a series of signals in sequence, a population of time-traces can be collected under similar conditions. Running a subtraction algorithm on these signals, since there can be assumed no component change, gives a measure of noise and variation.

For an absolute difference threshold T, first arrival signal amplitude of 1, and absolute noise level of N, the maximum acceptable path distance can be calculated:

$$T = \alpha \times N \tag{35}$$

$$T = \frac{Nd}{2} \tag{36}$$

$$d = \frac{2T}{N} \tag{37}$$

Using example values of $T = -25\ dB$ and $N = -40\ dB$ (both relative to the same reference), this gives a maximum distance of $\sim 11 \times transducer\ seperation$.

3.5.4. Background electrical noise

Electrical noise is the addition of unwanted contributions into information-carrying signals due to the coupling of circuits with other electrical signals.

In many cases, electrical noise is random, with frequency outside the bandwidth of interest. In these cases, if the ratio of signal amplitude and noise amplitude is high, averaging and frequency filtering in hardware and software can be used to suppress this noise so that its contribution is of little significance. Some contributions, however, are not random, or are high amplitude compared to the signal of interest.

It is useful to have a method for determining approximate electrical (background) noise level, to give confidence that it will not affect the overall operation of the system. In order to

determine an order of magnitude effect, noise can be extracted from a recording without any signal of interest. This can be achieved by measuring without transmitting a pulse. Alternatively, if data has already been collected, as is the case in this work, by using data collected at the beginning of recording time, before the pulse emission (i.e. t < 0). This background noise signal can be extracted, and an RMS measurement performed, to understand electrical background noise.

For the example dataset, taking a section of signal before the pulse emission time, the maximum amplitude compared to the signal maximum amplitude was -43 dB. This is significantly smaller than the detection performance desired (-25 dB was typically used in this work).

Later, the impact of background effects on performance will be shown in Figure 9. This is performed by comparison of the difference between 500 datasets recorded in sequence (at the same conditions and system age, etc.)

The effect of this may be greater when considering transducers such as EMATs, or in high-electrical noise conditions, such as when the measurement system is in close proximity to high-voltage electrical equipment (e.g. motors and generators). Hardware design and techniques such as physical shielding are vital in these applications to ensure usefulness of gathered data. Background electrical noise is known to be substantial in some cases – especially when manufacturing activities, such as welding, are being undertaken.

3.6. Difference calculation

The chosen method works based on the ability to establish a match of signals through signal comparison. There are a variety of methods that may be used to compare two signals.

The method used in this work involves, first, the subtraction of one signal from the other, resulting in a "residual" signal equal in length (number of data points) to the original signals. For a perfect match of signals (i.e. if subtracting identical signals), the result would be an array of zeros.

There are two methods for how this subtraction can be performed. Subtraction of RF signals, or subtraction of the Hilbert envelope of each signal (dotted and solid blue lines of Figure 3, respectively).

Envelope subtraction has been shown to give an improved signal to noise ratio. The reduction in noise results from the slower rate of change of the envelope signal compared to

an RF signal and therefore, small time offsets have a smaller effect. However, this comes at the expense of introducing blind spots into coverage (Croxford, et al., 2007).

Since these small time-offsets can be accounted for by shift operations, and because of the importance of full coverage in the monitored area, RF subtraction was selected.

The second part of the process is to take the residual signal calculated, and convert it into a single value metric that describes its size, and so can represent the overall difference between the two original signals.

The two options considered in this work for single-value quantification of a difference metric are: maximum (absolute) value of the residual, or RMS (root mean square) value for the residual.

The choice of optimal metric is dependent on the specific application, and response of target defects. A target defect in a relatively feature-sparse application will present as an isolated reflector, giving a clear individual peak. This is suited to the use of a metric such as the maximum of residual, where a single change in peak amplitude indicates the presence and size of a defect – similar to the approach of conventional UT interpretation. The other case is in feature-dense applications, where a measured signal has many benign reflectors, and where a similar target defect will give many reflected responses from different paths. In this case, it is more appropriate to look at the sum of all changes due to the presence of a new reflector – e.g. RMS of the residual signal.

It should be noted that whatever measure or metric is calculated to describe the similarity between two signals, this does not provide any accurate measure of defect size or characterisation. The metric acts as an indicator and tracks with defect size. Proper NDE validation activities must be undertaken to identify critical defect sizes and morphologies, as well as justify probability of detection. In the absence of this, measurement of difference between two signals under varying condition, is the best generic measure of performance.

The metric that used the maximum-of-the-residual was selected for the majority of this work, as the focus was the detection of the simplest case of individual high amplitude peak reflections.

A simple comparison of each metric was performed on a dataset of similar signals. Figure 9 shows the spread of difference-metric values obtained from 500 signals, captured sequentially, under the same conditions. The two plots in Figure 9 provide a comparison of the maximum of the residual signal (relative to the maximum of the baseline signal), against the root mean square (RMS) of the residual signal (relative to the RMS of the baseline signal).

For work that required a threshold to be set for when two signals were to be considered a match, a metric of -25 dB was used (for maximum of residual compared to first arrival signal amplitude). This provides a low-false positive rate based on the results presented in Figure 9.
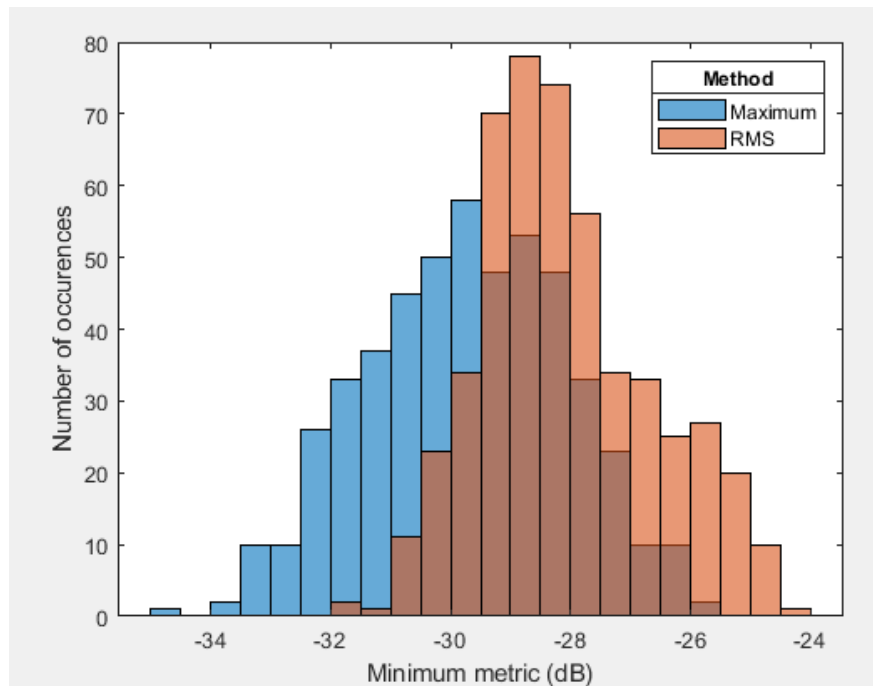
Figure 9 Histogram of the minimum calculated difference metrics for 500 data sets collected sequentially under the same conditions.

The results show both methods produce (approximately) a normal distribution of difference values for these 500 data sets, with a mean metric value slightly differing between the two metrics (-30 dB for maximum, and -28 dB for RMS). This investigation assumed that no change occurred over the time of recording, such as heating of the instrument through prolonged use.

The spread observed is similar between both metrics, of approximately 8 dB, giving confidence in the similarity of behaviour under these conditions.

Broadly, it can be assumed, then, that an approximate correction of around 2 dB may be used between each metric where signal changes match that of these variations under similar conditions.

This is not, however, valid for all cases. Depending on defect response – e.g. due to number of features reflecting the damage and so number of responses – the effect on RMS and maximum of residual will be different. It is therefore important to undertake case-by-case assessment of metric for the application and defects of interest.

It is proposed that future work could be investigation of a hybrid of these methods. Due to differing applicability of each method described above, that a combination metric (calculated from both the RMS and maximum) may give coverage of both cases, and lead to a more adaptable monitoring system.

3.7. Shift operation parameters

The operations described in the 'baseline stretch' section of the training algorithm modify the signal in several different ways in order to attempt to accommodate known (benign) physical variations. The extent of each of these stretching or shifting transformations is varied over a certain range, for example, the signal may be shifted in time between a minimum and maximum shift values, with a step size between each trialled value.

Using too many parameters in a test grid would make it time consuming, with some never being used. For the complete dataset used in this work (5000 signals, of 4000 data points each), for example, a full computation of shift operation may take, in the order of, 24 hours on a mid-performance machine. There are several methods for determining the range, and coarseness of the parameter range, which shall be explored. However, it should be noted that it is not possible to find and set a single parameter range that is suitable for all applications, or even all data sets. Instead, it must be calculated every time a new data is used. It is, therefore, useful to have a robust method of calculating this prior to full data processing, allowing the training process time to be reduced to a practical amount.

Prior knowledge can be used to provide a good initial guess of a parameter range, useful for an initial guess, before further optimisation can be carried out. Environmental and operational conditions will be understood to vary within certain reasonable bounds. Therefore, the effects on the data can be assumed to be limited to certain predictable ranges. Through this, a reasonable theoretical upper and lower limit can be determined, also considering that the assumptions of Baseline Stretch Subtraction only tend to be valid for small shifts.

This range, however, may still be surplus to that required to find an appropriate solution. By initially running through a number of the datasets with a single baseline (restricting baseline growth) an understanding can be gained of the maximum range of each parameter used, outside of which is wasted processing.

A valuable property to evaluate when investigating these shift/stretch operations, and their ability to compensate for EOC effects in the data, is to determine whether they are dependent. A dependence, or coupling, of multiple shift operations means that the operations and/or EOCs that they are representing, are interacting. If the shift operations are determined

to be independent, this permits assumptions to be made that can significantly reduce computational burden, and simplify the prediction of system behaviour under new conditions. For example, an optimum stretch value for one operation could be determined independently of the others, reducing the total number of combinations of stretch operation parameters that need to be calculated - in this example, the parameter space to be tested is reduced by one dimension.

3.7.1. Time translation

Time translation involves the addition or subtraction of a constant value to the time-base. This operation is used to correct for the jitter shift. Jitter shift occurs in the instrument, changing the position of pulse emission about $time = 0$ by a small amount.

Considering the shifting of a time-domain signal $F(t)$ by a time δt, in order to permit non-integer shift values, first the signal is transformed into the frequency domain using the fast Fourier transform (FFT) for computational efficiency. The frequency domain signal, $F(f)$, is then then shifted using the operation:

$$F(t + \delta t) = ifft(F(f)e^{-i2\pi f\, \delta t}) \tag{38}$$

where f is the frequency vector.

The operation implementation permits non-integer values of time translation to be used. It was believed, however, that on the dataset considered, this effect is simply an integer translation in time.

In order to evaluate this, a test was performed on 1000 signals (recorded over several years) using an equally spaced array of time translation values between -5 and 5 time points, using steps of 0.01 time points.

Figure 10 shows a histogram of time translation values that gave an optimum match between datasets and baseline.
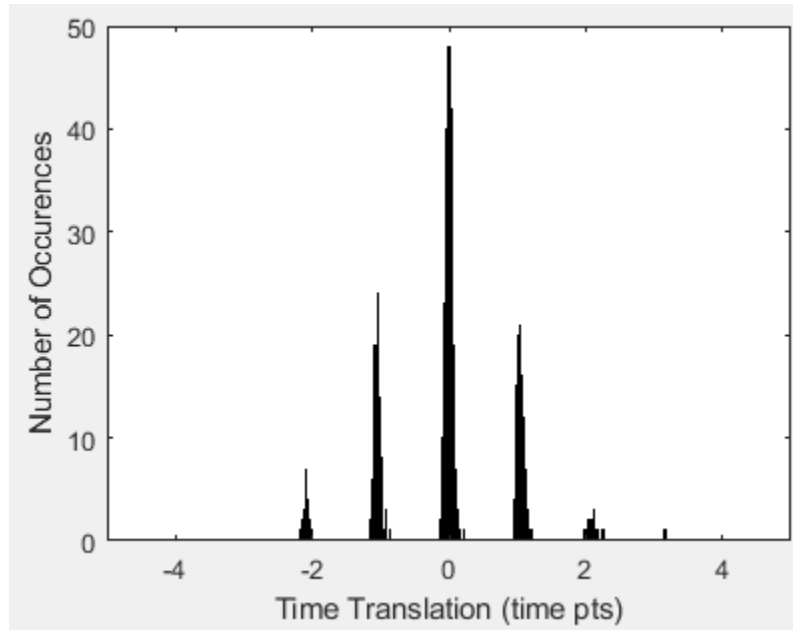
Figure 10 Histogram of time translation values used to find the best match between 1000 datasets

It was determined that nearly always, integer values of time translation were used, showing that the instrument translation was in fact a single time point jitter effect. Additionally, no values above ±2 time points were used.

It was therefore proposed that a suitable range of shift values would be -2 to 2 time points in steps of 1 time point. This would accommodate for a possible deviation of ±1 from a median value, with a -1 (from the median) signal needing to be shifted +2 to match a +1 (from the median) signal, and vice-versa.

This result adds validity to the assumption that time translation may be considered independent of other shift values, as it is due to an effect of the recording instrument that simply changes the point of signal emission by ±1 time point. Additionally, the effect is observed most significantly on the first arrival signal. Because of these two factors, it is therefore possible to separate the jitter shift (time translation) operation from all other operations, reducing the processing time by a factor (approximately) equal to the number of shift values used (5x).

To verify the effect of reducing the number of trialled shift values in the parameter range, the training process was run again with the proposed reduced set.

Figure 11 shows the small deviation in performance between using the full range of time translation values, and using only integer values between -2 to 2.
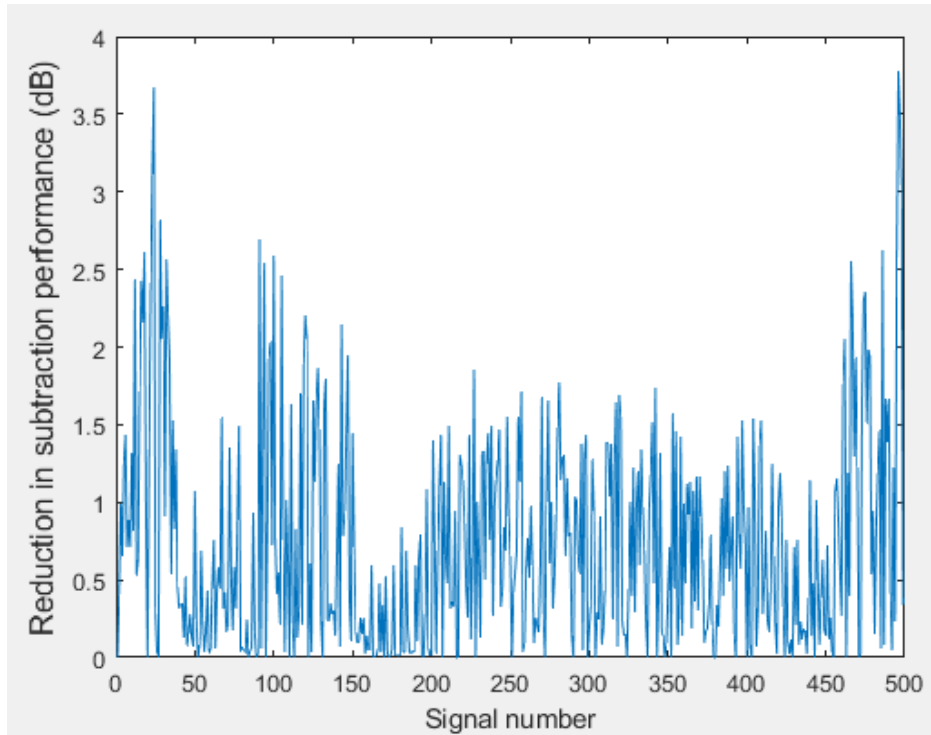
Figure 11 Difference in signal subtration performance with a single baseline of using the proposed reduced time translation parameter range and a full parameter range

It can be seen that the performance decrease from only using five time-translation values (compared to 121 in the fine array) was on average 0.7 dB.

A valuable property to evaluate when investigating these shift/stretch operations, and their ability to compensate for EOC effects in the data, is to determine whether they are dependent.

To determine whether effects are independent, an assumption could be made that the same optimum compensation for one operation will be reached, regardless of whether other compensation operations are performed. That is, the same optimum time translation value would be used if performed in isolation, or if a global optimal solution is found with time translation, time stretch, amplitude scaling, and phase shift.

A test was set up in which the first case varied only time translation to find an optimum match for each of 500 signals. The second, time translation, phase shift and amplitude were also varied to find a global best match. A single baseline was used, so that differences in baseline growth did not affect the results. The time translation values were compared, and 83% of results matched between the two tests. This was not as high as expected and indicated that the assumption of independence of time translation was inaccurate.

It was, however, considered that another effect may have impacted the results. In the cases where it was apparent that optimum time translation value was affected by other parameters, this may be due to a limitation in the time stretch model used. Due to the use of a single baseline, the time stretch operation may have been inappropriate for some of the larger temperature variations. The time stretch temperature compensation assumption is only valid for small temperature changes, as the true effect of temperature is to shift the arrival time of signals, rather than effectively stretch the while time-base. In this case then, time translation may have been a better operation to compensate for these larger temperature shifts. As the actual technique used, uses a combination of OBS and BSS, these larger temperature variations would be better matched by new baselines at higher temperature.

Instead, a better measure may be the subtraction performance when the time translation is used in conjunction with other shift operations (assume dependence with other operations), and when the optimum time shift is calculated independently of other operations.
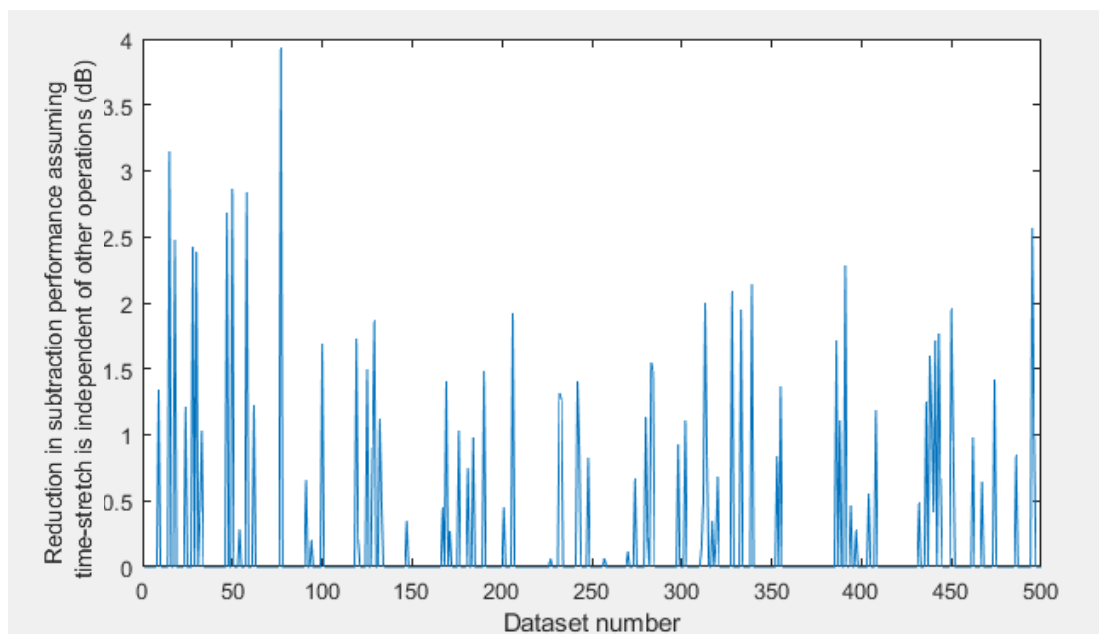
Figure 12 Subtraction performance with time translation treated independent or dependent of all other stretching operations, with a single baseline

Using a single baseline, the mean reduction in signal subtraction performance (measured by the metric: maximum amplitude of residual) when the optimum time translation search was performed independent of any other stretching parameters (amplitude scaling, time stretching, phase shifting) was -0.2 dB (Figure 12).

The above results indicate there is some coupling between time translation and one or more other operations. It was proposed that the only operation that optimum time translation is

dependent on is amplitude, since amplitude is so closely linked to the difference-measure used. Previously, the results were presented of performance when optimum time translation was determined independently of any other operations. An additional test was performed where optimum time translation was found, whilst varying amplitude, then this time translation value was used to find the optimum time stretch, amplitude scaling and phase shift combination. There was no performance improvement over optimising time translation without amplitude variation.

Although potentially useful to consider for other applications, the performance improvement on this dataset was not significant.

3.7.2. Amplitude scaling

Amplitude scaling is used to account for the known effects of temperature on transducer response, and random variations in the measurement system.

Amplitude scaling involves multiplying all amplitude values in an A-scan by a single value. The centre of the signal must be around amplitude = 0, that is, there must not be a DC offset.

A reasonable pair of limits for amplitude scaling can be determined by the minimum and maximum amplitudes of the first arrival signal. This is a computationally inexpensive method for providing an initial guess that can be improved and iterated on.

The (absolute) maximum of each of 1500 signals was taken, before pre-processing, corresponding to the first arrival signal amplitude. The lowest first arrival amplitude was 0.88 of the highest. Taking a conservative estimate, a maximum required amplitude scaling of 15% was determined.

An array of amplitude scale values can then be defined with the step being determined based on required sensitivity. Considering a step size ΔS where the true solution lies exactly between the two nearest points of the parameter space (i.e. the worse-case), this maximum error of amplitude scaling is equal to half the step size. This inaccurate scaling value would result in an incorrect scaling of the signal. Therefore, the maximum error from using this parameter space for an amplitude A is given by:

$$error = \pm A \frac{\Delta S}{2} \tag{39}$$

Considering the largest signal (first arrival signal normalised to amplitude $A = 1$), for which the error is greatest, and a difference threshold of -25 dB, which sets the sensitivity. The error in amplitude scaling that would cause the first arrival signal to break the threshold is given by:

$$20log_{10}(error) = -25 \tag{40}$$

$$error = 10^{-\frac{25}{20}} = 0.056 \tag{41}$$

$$\frac{\Delta S}{2} = 0.056 \tag{42}$$

$$\Delta S = 0.11 \tag{43}$$

As this is the maximum step size that would not cause the difference metric to be reached erroneously, a conservative value of $\sim \frac{1}{4}$ was selected, of $\Delta S = 0.003$.
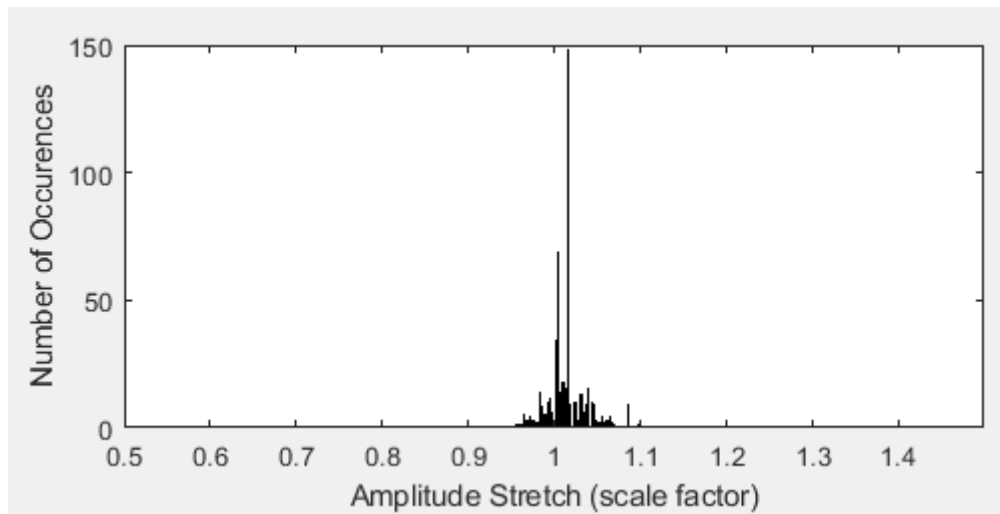
Figure 13 Histogram of amplitude scale values used to find optimum match between (growing) baseline set and 500 signals.

Figure 13 shows a tight range of optimum amplitude values used in the matching of signals to baselines. This illustrates the suitability of the range of scaling values calculated earlier (15%), given a much larger parameter space trialled.

A test was performed to evaluate independence of this amplitude scaling operation. One baseline was used, varying amplitude only (so that it is independent of other operations), then comparing optimum amplitude scaling factors used with those used when all parameters were varied to find a global optimum set of stretch/shift values. The mean absolute difference in amplitude scaling value was determined to be 0.13. This is close to the maximum optimum shift observed (Figure 14). Therefore, it can be assumed that this test has demonstrated that there is very little similarity between the behaviour of the amplitude scaling operation when varying other operations. Amplitude scaling is the operation most closely linked to metric, and based on the results, it is believed this parameter should not be considered independent.
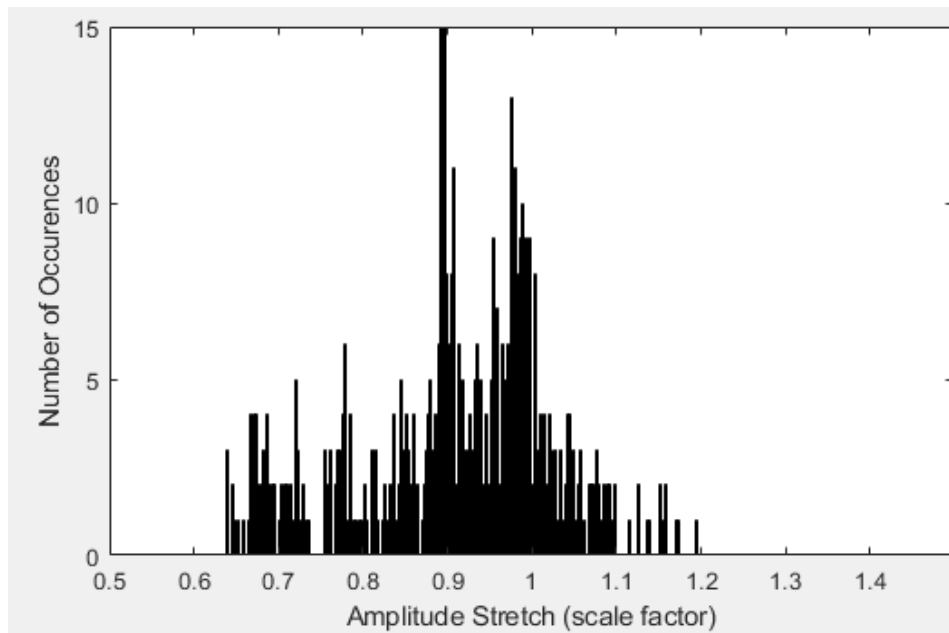
Figure 14 Histogram showing range of amplitude used to find optimum match when one baseline is used, and only amplitude scaling is varied (i.e., all other shift/stretch operations are not performed)

3.7.3. Time stretching

The time stretching operation involves stretching of the time axis of the signal. Strictly this is not an accurate representation of temperature effects on guided wave signals. The effect of temperature as a time stretch can be described as each received signal being translated in time, proportionately to distance travelled (Croxford, et al., 2007). This requires signal contribution from each wave packet to be distinguished and separately time shifted based on the centre of the wave packet. Since this is impractical, a stretching of the time base is used, which is valid for small shift values (i.e. for small temperature compensation).

Time domain stretching requires the resampling of data points, and extending of a dataset to create extra data points, for maintaining a constant time base when compressing a signal. This is performed using the scale transform, as described in (Harley & Moura, 2012).

The method for performing this operation correctly depends on the signal being continuous, requiring both the first and last time points of the signal to have an amplitude equal to zero – and for this transition to be gradual and smooth. This is performed by the cropping process detailed in Section 3.5.2 (Digital Filtering).A test was performed to determine the dependence of time scaling on other shift operations. The difference between optimum shift values was measured for when all parameters were varied to find an optimum global solution, and when only time stretch was varied. A single baseline was used.

Figure 15 shows, for each given dataset, the different optimum time stretch value when found in isolation, or with global varying of other operation variables. This shows the extent to which time stretch was dependent on time shift and amplitude scaling.

Figure 15 Difference in optimum time stretch value when found in isolation (independent of other operations), or with global varying of other operation variables (dependent on other operations).

The magnitude of the difference varied between three central values: 0, -470×10^{-6} and -940×10^{-6}. This can be explained by the absence of jitter correction (through time translation) in the case where optimum time stretch was calculated independently of time translation. It can therefore be said that time stretch is coupled with time translation.

To validate whether the optimum time stretch selected is predominantly coupled with time translation as hypothesised (i.e., not amplitude scaling), the optimum time stetch values were compared when only time stretch and time translation were varied freely together, with the test in which all operations were performed. Figure 16 shows the significant reduction in difference of optimum time stretch value (mean of absolute values of 2.8×10^{-6} compared to 420×10^{-6} in Figure 15).
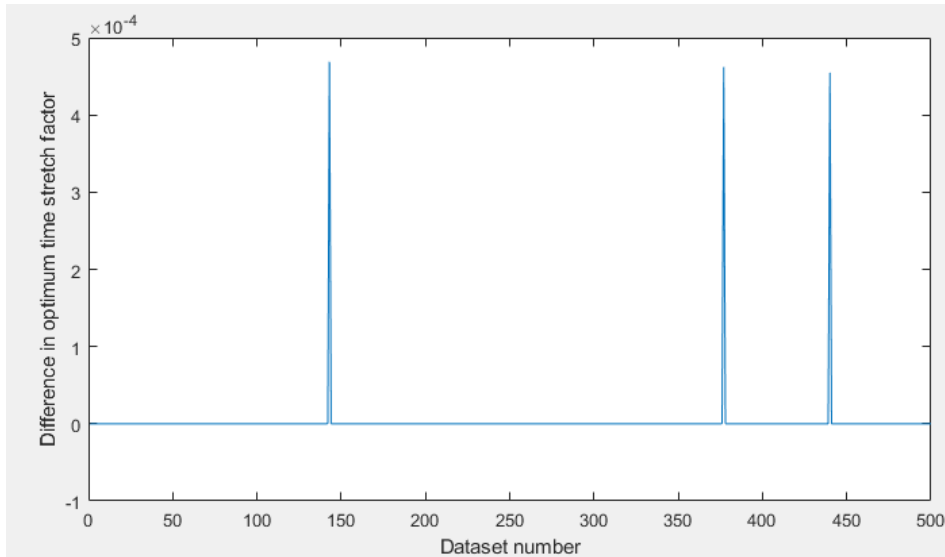
Figure 16 Difference in optimum time stretch value when found in conjunction with time translation (independent of amplitude scaling), or with global varying of other operation variables (time stretch, time translation and amplitude scaling).

Therefore, it was determined that the optimum time translation selected is coupled with jitter shift. A proposed improvement over both options – to give lower computational burden than calculating entire parameter space, but improved subtraction performance against treating time stretch as independent – was the optimum time shifting (jitter shift correction) of the data initially, before then determining optimum time translation independently of amplitude scaling. Figure 17 shows the results.
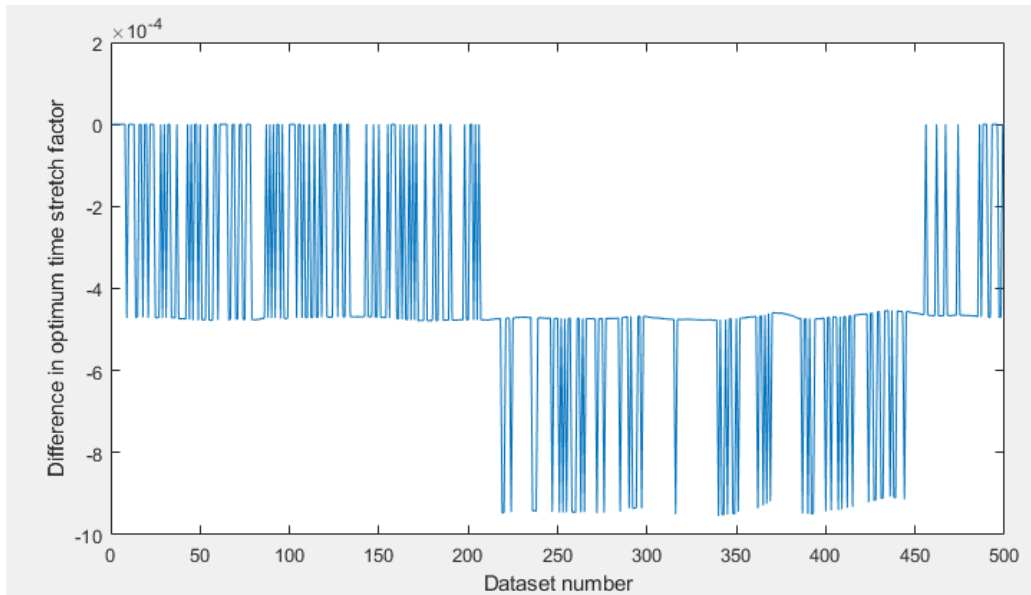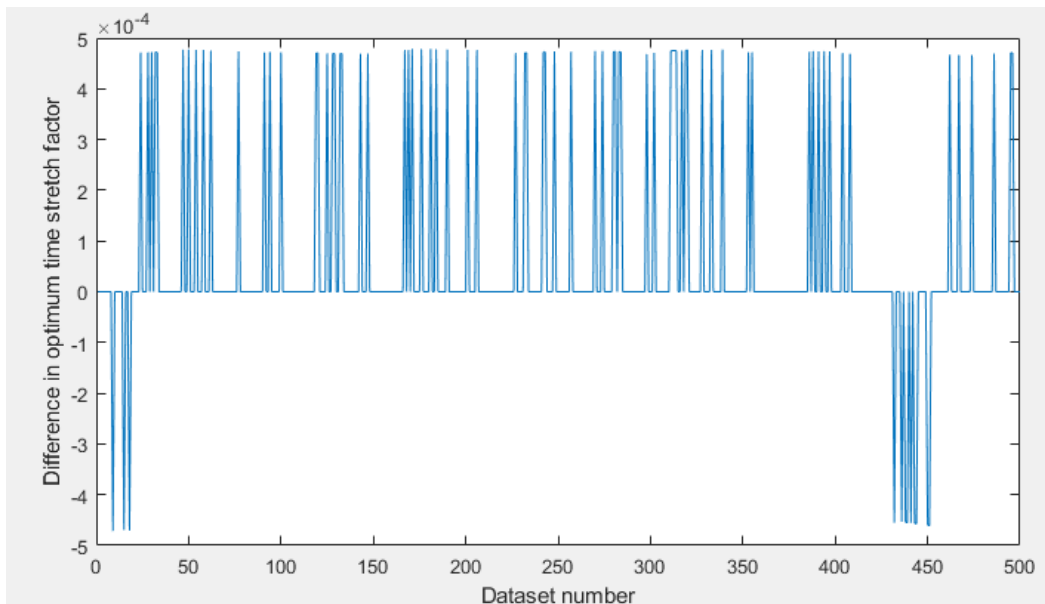

Figure 17 Difference in optimum time stretch value when found in isolation (independent of other operations), or with global varying of other operation variables (dependent on other operations). First, time translation (jitter shift was corrected), independently of all other operations.

The mean (of absolute) difference was 79×10^{-6} using this approach. This represents an improvement over the purely independent determination of time translation. The improvement only represents a five times increase in cases matching the full parameter case, still significantly different to the match found using the approach in Figure 16. This approach does, however, represent a greatly decreased computational burden.

This difference may also be due to the expected effect discussed in Section 3.7.1 (Time translation), where the use of a single baseline may lead to jitter shift correction being used to accommodate for large temperature differences that do not fit the time stretch model well.

To evaluate the time stretch values used to find an optimum match between signals at different times (and environmental conditions), Figure 18 shows the optimum time stretch values when a single baseline was compared to 4000 signals, with all shift operations performed (global minimum found).
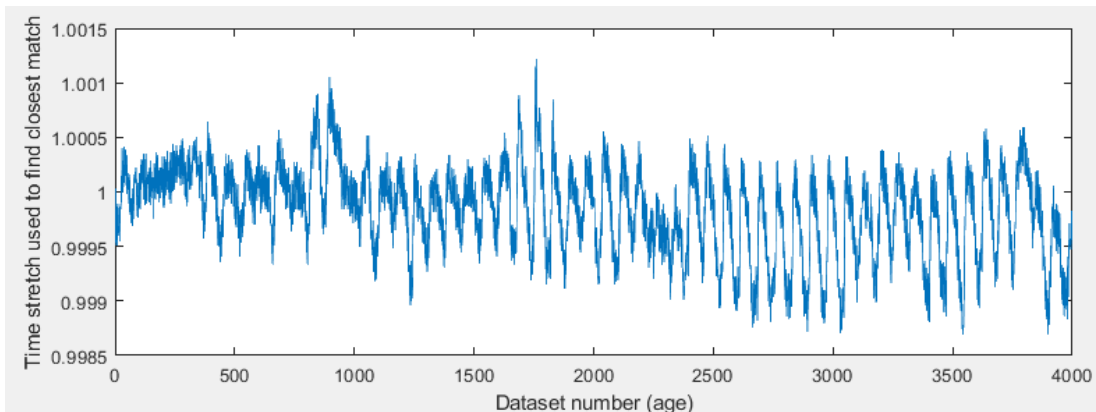
Figure 18 Time stretch (scaling) values used to find closest match between datasets and a single baseline

Temperature cycles can be seen in the data. These are thought to correspond with time of day at which each of the measurements was made.

When using single baseline, the maximum time stretch used was 1.0006, minimum 0.9994. Permitting baseline growth where the optimum match produced a metric above the threshold (-25 dB of the first arrival signal), the minimum and maximum time stretch values were 1.0003 and 0.9997.

This demonstrates a decrease in the use of time stretching to compensate for temperature change when new baselines are able to be added, which more closely match the temperature effected signal. This aligns with the understanding that the inaccuracy of the time-stretch/temperature model at large values of temperature change.

3.7.4. Phase shifting

A potential change of signal phase is credible, where aging and temperature variations may cause transducer behaviour to change between measurements.

Phase shifting of a signal is performed by multiplying the frequency domain representation of the signal by a constant value. For a signal $F(\omega)$, and phase shift $\delta\omega$:

$$F(\omega + \delta\omega) = F(\omega) \times e^{-i\delta\omega} \tag{44}$$

A test was run with 500 data sets, restricting baseline growth so that only a single baseline is used. A range of phase shift values were used between -2π and 2π $rads$ with a step size of $\frac{\pi}{100}$ $rads$, giving 401 total trialled values. By recording shift values used when the closest match was made to the single baseline, it was possible to identify whether a shift of phase ever improved signal matching. This provides the ability to assess the performance benefit of phase shift compensation against the processing speed penalty from increasing the parameter space.

It was determined that the vast majority of baseline matches were made with no phase shift, and the remaining low quantity had only a very small shift, as shown in Figure 19.
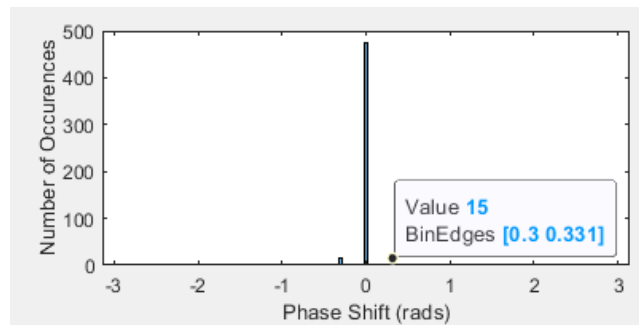
Figure 19 Histogram with bin widths of $\frac{\pi}{100}$ of phase shift values used to find closest match to a single baseline, with no other shifting operations performed.

Since no phase shifting was used in determining optimum match to a baseline, dependence of phase shifting on other operation could not be determined.

In other cases, this phase shift may be coupled with a secondary effect that cannot be compensated for using the operations described. It may be the case then, that despite a phase shift improving performance when using a single dataset, a better match can still be obtained by using multiple baselines, even without compensation for phase shift. In this case, training data must be sufficiently complete that it covers all EOCs that introduce the phase (and secondary effect) variation.

3.8. Speed optimisation

3.8.1. Overview

The processing of datasets as part of training is time consuming. The large number of combinations of stretch operations that need to be computed to find an optimal solution, combined with the large size of typical data sets, and the brute-force nature of the approach, mean significant computational burden is unavoidable. However, some effort was devoted to optimisation of the process that improve the practicality of data processing on hardware available for the project. As this project is not focused on the computation aspect of this challenge, proper optimisation and software writing are given as possible further improvements.

3.8.2. Use of the first adequate match

If a solution is found that is below the threshold, then it is known that the data sets are similar. No additional benefit (outside of experimentation work) is gained from knowing how similar they are. That is, it does not matter how "defect-free" the component is, provided the difference metric is below the threshold. Therefore, as soon as a subtraction is found that gives a residual below threshold, the loop can be broken, and the data set classed as similar, ready to move on to the next data set. This is shown diagrammatically in the bottom section of the flow diagram in Figure 20.

It should be noted that this was not used in the investigations above, as there was significant value in determining the overall minimum when developing and evaluating the system performance.

3.8.3. Ordering of operations

The order that each of the operations are performed when calculating the parameter space can affect performance. In addition to the obvious computation savings, such as having the most computationally heavy operations on the outermost programme loops as possible (so they are performed a lower number of times), other time-savings were proposed based on the objective of algorithm.

The reordering of parameter ranges was performed to first trial the smallest shift values (e.g., order 0, 1, -1, 2, -2), as it is expected that small shifts are most likely to give matches. In conjunction with loop breaks when a solution is found, this increases the likelihood of a solution being identified earlier in the total parameter-space search.

3.8.4. Baseline sorting

With the breakpoints described in Section 3.8.2, it now becomes beneficial to sort the baselines in a way that results in a match being identified as early as possible.

Since the time-of-year dependent effects have a low frequency of change, it can be assumed that the most likely match for a given signal will occur in a baseline that was recorded recently (i.e., at a similar recording time/date to the signal). Therefore, it is proposed that reverse ordering baselines, so that the most recently added entries are trialled first, will result in a match being found in less time.

It may be proposed that a match may also be likely for similar times-of-year for previous years. However, the age dependent effects apparent in a provisional trial of the data (discussed in more detail in the conclusions of Section 3.9), result in this not being the case. Therefore, it is proposed that the method above is sufficient to give a performance improvement from baseline sorting.

It is noted that baseline ordering (in conjunction with loop break) will result in the full parameter space not being tested, and so will greatly affect and skew which baselines are used. Therefore, any investigative work where the best-matched baseline overall is of interest must have these loop breaks (detailed in Section 3.8.2) disabled so that the entire parameter space is tested, and the true global best match found.

3.8.5. Overall Training Algorithm

Figure 20 shows the changes made to the training algorithm, based on the investigation into performance on a specific dataset.
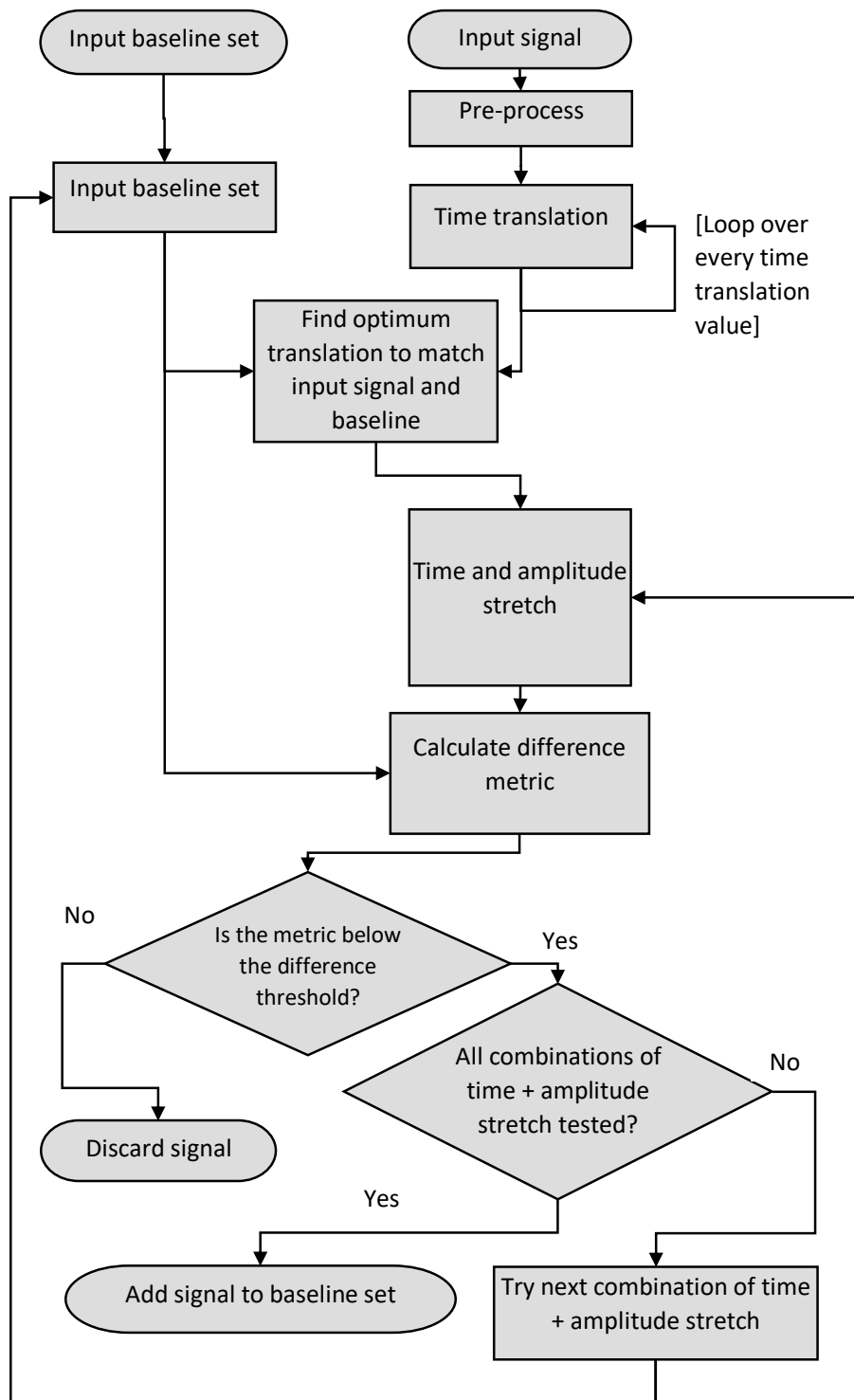
Figure 20 Training algorithm flow diagram, with computational savings and time-translation operation performed independently

3.9. Overall results

Each of the preceding sections has dealt with performance improvements from individual changes to the subtraction process. Each of these changes offers an option to tailor the system to choose between processing speed and model behaviour, with overall subtraction performance and suitability for incomplete datasets.

To understand the overall performance difference in each extreme of these cases, a comparison was made between a test where all parameters were permitted to vary freely and interact, with another test where the behaviour was restricted, and assumptions of independence between various shift operations were implemented to reduce the search for an optimal match to only the most likely combinations.

Table 1 shows the performance comparison (subtraction and time) between the two approaches (best subtraction performance vs with computational savings) for a total of 500 data sets of 4000 time points each. This was performed on a mid-performance desktop machine (with an AMD Ryzen 5 3600, 6-Core, 3.6 GHz Processor).

Table 1 Subtraction performance comparison of two implementations of the subtraction algorithm, with differing levels of restriction on combinations of shift operation parameters and dependencies

	Implementation	
	Best subtraction performance	**Reduced computation**
Subtraction performance (one baseline)	-18.4 dB	-17.4 dB
Computation time (one baseline)	~5 hours	~30 mins

This was performed with a single baseline, with the performance difference shown above being conservative for the actual OBS+BSS method where baselines are added. Since additional baselines will result in smaller shift operation ranges needing to be used, as a more representative, accurate (non-stretched) baselines to any new signal will exist, and since smaller shift operations typically resulting more accurate modelling of EOCs, this will always improve performance over the use of a single baseline. Therefore, the practical benefits of using the entire parameter space can be assumed to be significantly reduced when using multiple baselines.

3.10. Conclusions

The algorithm used for this work employs various compensation operations that model the behaviours of instrument and environmental effects on recorded signals. These models are imperfect, and have limitations in the extent to which they closely follow observed behaviour. The multi-modal signal contributions, and complex interactions with features of the structure, reduce performance of each of these compensation models. However, by permitting the free variation and interaction of these operation parameters to find a global minimum, it is possible to extend the performance past that of simply modelling each change independently. In this case, the role of compensation of each of these operations has been observed to overlap, providing an improvement in subtraction performance. An example of this in the work described, was the interaction of optimal time-stretch value when a jitter-shift (time translation) search was or was not implemented. When there was no time translations, the time stretch values used varied between three discrete nominal values, indicating that some time translation of a single time point was being performed using a time-stretch.

The best overall performance can be achieved when the search variables are allowed to vary freely over the entire parameter space. However, with a slight reduction in subtraction performance, a significant time saving can be made, accompanied by a control of the role of each shifting operation. This improvement in understanding of behaviour is beneficial for validation and verification. An example of one of these controls is: when time translation is calculated independently of other shifts, and without a DAC applied, it is used as a jitter shift correction only. This prevents time translation being used by the algorithm to compensate for poor time-base stretching compensation for large temperature differences between two signals.

Previous work on the same dataset found that (maximum) signal amplitude, centre frequency and bandwidth varied with changes in temperature, but did not exhibit any long-term trend over the duration of recording (Courtier, 2018). Therefore, it can be considered that these parameters are not ageing effects in this case.

There is a considerable long-term change in signals over the monitoring period, observed as the low frequency of a given recorded signal matching with baselines that was recorded much earlier in the monitoring period. It is beneficial to establish whether the measured change over the recording period (of three years) are related to sensor systems aging effects, or are in fact due to structural changes (which may indicate real structure degradation).

4. Accelerated Aging Trial

4.1. Overview

The previous section indicated an age-dependent effect that imposed a limitation on the performance of the installed sensor system. It is believed that this effect may be due to changes in the sensor system that do not correspond to damage or degradation of the structure.

The existing results make it difficult to separate bond/transducer aging from structural changes. Therefore, it was believed beneficial to undertake a trial where it was known that no test specimen change occurred, conditions could be controlled, and where repeats could be conducted to provide more confidence in reliability of results.

To try to determine whether any of this noise can be removed by understanding the age-dependent effects of bonding on recorded signals, a trial was devised. Accelerated aging of a typical transducer and bond was attempted, with long-term aging simulated in a much shorter space of time through high temperature exposure and temperature cycling.

In order to achieve this, first a hardware system needed to be set up.

4.2. Hardware system

For this work, 1 mm thick, 20 mm diameter PZT (lead zirconate titanate) piezoelectric transducers were used, which were optimised for 50-200 kHz in a plate thickness up to 3 mm. A single pair of transducers was used per sample, operating in pitch-catch modality. These were bonded to 2.5 mm aluminium sheet, of approximately 300 x 100 mm. The bonding method used was a two-part epoxy adhesive (Araldite (Araldite, 2011)).

The bond was achieved by applying a slight excess of adhesive onto the mechanically abraded and solvent cleaned surface of the plate, positioning the unsoldered ceramic transducer, then pressing firmly to remove any excess adhesive between the transducer and plate. A consistent mass of approximately 1 kg was then applied for 5 minutes until the bond was set. A further 24 hours were given to ensure consistent curing of the bonds across the set of test samples.

Four samples were used, three of which were temperature cycled to provide an understanding of variation and confidence in repeatability, and an additional control was used as a reference from which to assess the significance of any changes.

A USB Oscilloscope with function generator (Handyscope HS5), was used for data acquisition, controlled by laptop. A 130 kHz high-pass hardware filter was used to remove any low frequency contribution from external vibration. A diagram of the setup is shown in Figure 21.
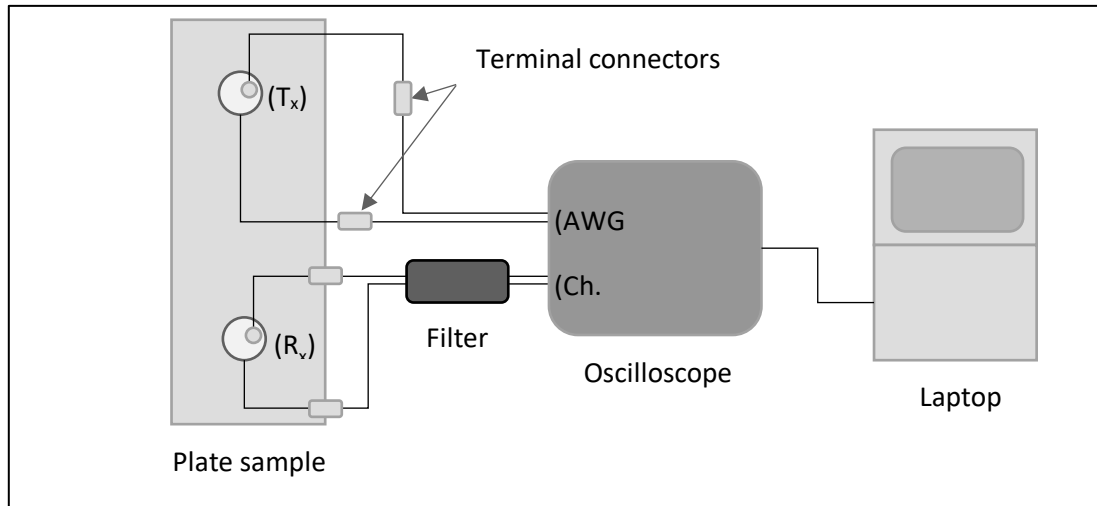

Figure 21 Diagram of hardware setup used for accelerated aging work

4.2.1. Velocity and Frequency Thickness

To be able to perform analysis of the recorded signals, it is necessary to determine the properties of the guided wave of interest. A primary component of this is the frequency and plate thickness dependent propagation velocity of lamb wave modes. Any other setup with the same product of frequency-thickness (as well as same material, etc.) will behave in the same way. This is useful for transferability/generalisation of results, and makes it possible to create laboratory setups with high frequency low thickness that emulate a realistic industrial application on high-thickness materials. The benefit of this for trials, is that the acquiring, safe handling, heat treating and storage of thinner plate than that used in marine applications makes it significantly more practical.

Wave packet speed is determined by using a dispersion curve diagram (like the one shown in Figure 22), for a specific value of frequency and plate thickness product. For a 250 kHz centre frequency input pulse in a 2.5 mm aluminium plate:

$$Frequency\ thickness = 250\ kHz \times 2.5\ mm = 625\ kHz\ mm \qquad (45)$$

Using a dispersion curve in Figure 22, approximate group velocities for zero-order (fundamental mode) anti-symmetric (A_0) and symmetric (S_0) modes were determined to be 3100 and 4750 m/s, respectively.
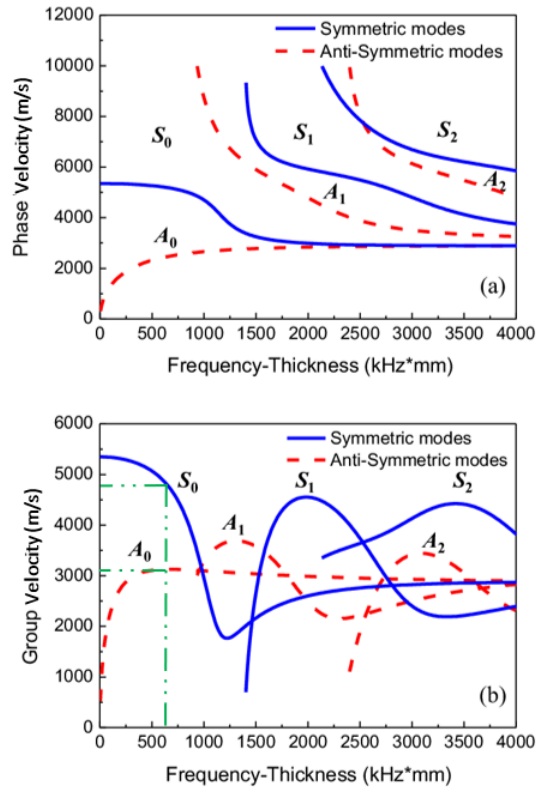
Figure 22 Aluminium dispersion curve (Zhao, et al., 2017) with values annotated (green) for a frequency thickness product of 625 kHz mm.

4.2.2. Acquisition parameters

A five-cycle Hanning tone-burst was used as an input signal in line with previous work in this area (Clarke, et al., 2010) (Croxford, et al., 2010).

Since acquisition time and data size were not of significant importance compared to other aspects of the experiment, a high sample rate, bit depth and gate length were used. The sampling rate was set at 5 MHz for a 200 kHz centre frequency, giving a large margin between the useful frequency range and the limiting Nyquist frequency. The use of 12-bit digitisation was sufficient for the test.

A gate length of 500 µs was used, which, using the slowest fundamental wave mode speed of 3100 m/s, gives a propagation distance of 1.55 m (five time the longest dimension of the plate).

4.2.3. Noise measurement

In order to predict the noise level, and therefore performance, of the overall monitoring system, a test was conducted. 500 recordings were collected in sequence, to cover a statistically significant representation of the variability of the recording system at a single operating point and age. The results were then passed through the training algorithm (described in Section 3), to give a quantitative indicator of noise suppression under ideal

48

conditions. The result of this is the absolute limit of certainty around any results produced using this method.
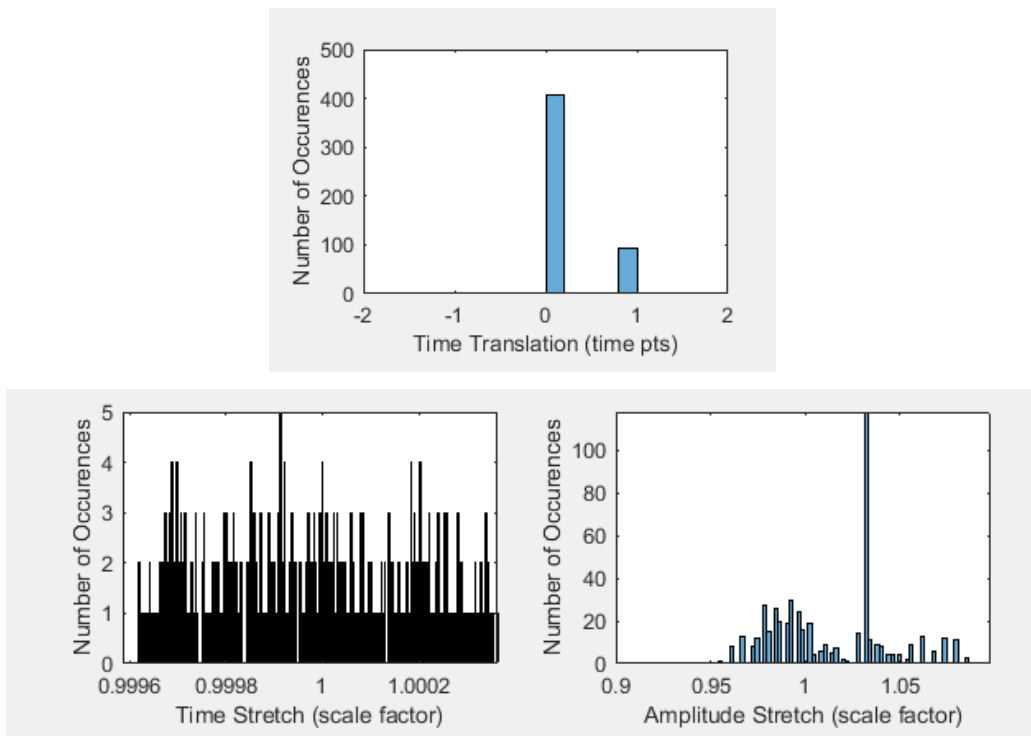

Figure 23 Parameters used for training on 500 signals recorded sequentially, under the same conditions

The tests show a low variability of stretch parameters used (Figure 23), with less than ±10% variability in amplitude scaling, which is observed to be the most variable feature of the signals. The variability of time stretch (temperature compensation) of less than 0.05%, verifying that spurious temperature compensation has not occurred. This corresponds to approximately a 3°C temperature change, based on an approximate fractional change in arrival time of the wave packets of 1 m/s per °C. The results demonstrate that the training is operating correctly when no compensation is required.

4.2.4.   Test conditions and recording regime

Four test pieces were used: one control samples, and three that would be cycled through high temperatures to simulate long-term aging effects.

Manufacturer guidance on maximum operating temperature for similar standard two-part epoxy glue state values of ~90°C for continued exposure, and 150°C for short periods (Gorilla Glue, 2019) (Permabond, 2014) (Araldite, 2014).

The samples and bonded transducers were left in an oven varying in temperature between 60°C and 100°C, before cooling to room temperature for data recording. All data acquisition was performed when the samples had reached ambient temperature.

A reading was made every 1 – 2 days (one cycle of temperature), for a total of 14 days to identify whether accelerated aging could be identified in the data. In general, the recording frequency of a monitoring system is set based on how rapidly operating/environmental conditions change, to capture data to represent all asset states sufficiently. For this test, all measurements were made at (broadly) the same conditions, so it was not necessary to have a recording frequency as high as a formal monitoring exercise. Instead, the recording frequency was set to capture trends over the timescale of the bond's aging process.

Instead of a single measurement being made at each age, ten recordings were made sequentially. The mean was then calculated at every time point for all ten recordings, giving an averaged signal. This averaging of multiple time-base data sets was performed to reduce any random noise component that may have introduced variability.

4.3. Results

4.3.1. Signal selection

The first arrival signal was isolated using the same time-domain cropping method described in Section 3.5.2 (Tukey window). The cropped first arrival signal for one of the three samples, as well as the control sample, is shown in Figure 24. No digital frequency filtering was performed.
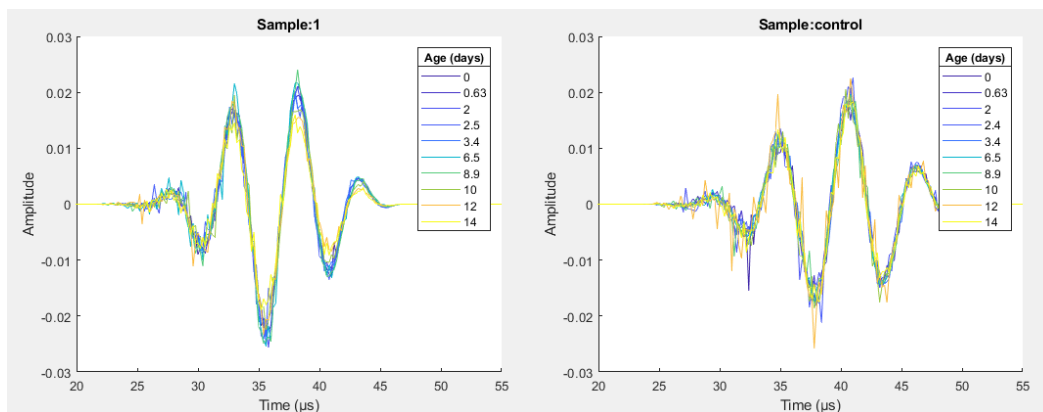
Figure 24 First arrival (time domain) signal from sample 1 (left) and the control sample (right)

The selected signal was then transformed into the frequency domain using the Fast Fourier Transform (FFT). The frequency content of the selected signal is presented in Figure 25.

As the first arrival signal constitutes only a small proportion of the overall time trace recorded, the number of samples was very low. To up-sample the frequency content for easier analysis, the time-domain signal was zero-padded after the signal of interest, before the FFT operation was performed.

Figure 25 Frequency content of windowed signal for sample 1 (left), and control sample right)

From an initial assessment of frequency response (Figure 25), it was apparent that the control sample had lower variation in amplitude of the centre frequency than the results from the heat-treated/aged sample.
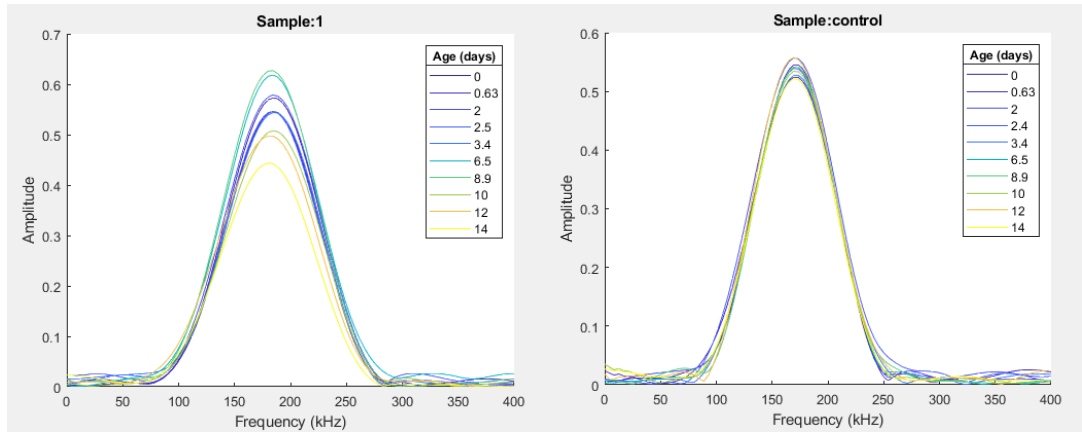
To quantify any changes, and permit comparison between signals, three measures were extracted from the frequency-domain content. The -6 dB bandwidth, centre frequency, and (maximum) amplitude of centre frequency, were used to assess the transducer response over the aging period.

4.3.2. Bandwidth

Bandwidth was calculated as the difference between the two frequency points on the frequency plot that have amplitude equal to half (-6 dB) of the maximum amplitude of the frequency plot (centre frequency).
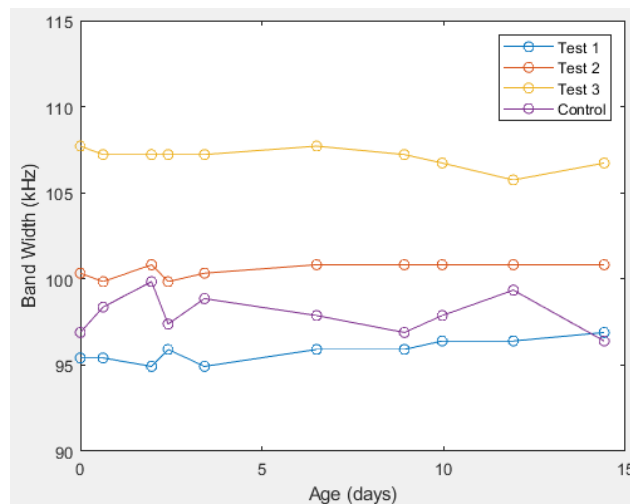

Figure 26 The -6 dB bandwidth of first arrival signal as a funciton of age for tests one to three, corresponding to the three samples, as well as the control sample that was not exposed to tmeperature cycling

51

The bandwidth variation (Figure 26) over the aging period shows no apparent trend. It is clear, however, that a certain amount of variability exists between each sample, much greater than that for each individual sample. This is expected to be due to differences in transducer bonding caused by inconsistency in methodology for attachment.

Since these data points represent the metric calculated for an average signal calculated from ten individual time traces, it was believed that additional information about variability may be gleaned from calculating the metric for each of the individual time traces before averaging.
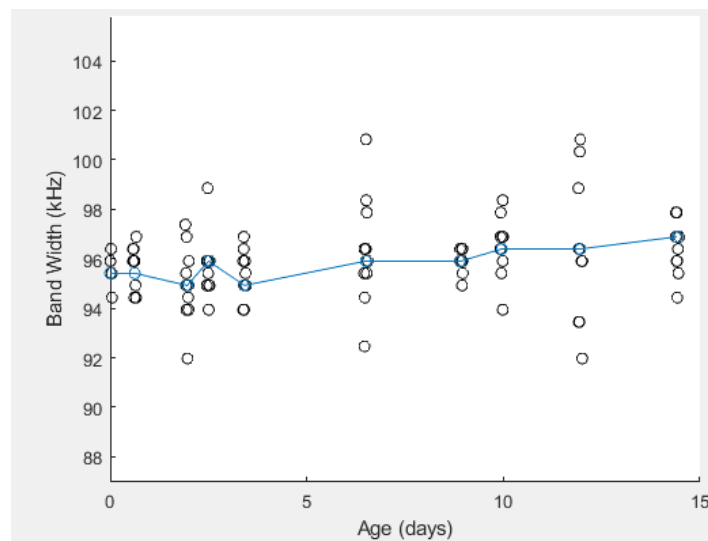

Figure 27 Bandwidth for sample 1, with black circles representing measurements for each time-trace, and the blue line representing the values for the averaged time trace. A small x-axis jitter is added to each data point to improve clarity where data points overlap.

Figure 27 shows the sample 1 values measured for each of the 10 individual recorded time trace at each age (before averaging time-domain signals), as well as value for the averaged signal. This figure represents results for the (similarly coloured) line for sample 1 in Figure 26. The results of the other samples are given in Figure 39, Figure 40 and Figure 41 in the appendix. It is worth noting that the data points in Figure 27 are given in a scatter plot, as there is no link between any single data point at one age, to another data point at another age.

Assuming there was no trend with age, the standard deviation of all measured values for sample 1 was calculated as 1.6 kHz, or ~2% of the mean, broadly consistent with the other samples (see Table 2). The variation in mean values, however, was much higher. The observed difference between samples is quantified in Table 2, showing a 11 kHz maximum difference in mean value between sample 1 and sample 3, equating to 11% of the mean of all values for all samples.

Bandwidth (kHz)		
Sample	Mean	Standard deviation
1	96	1.6
2	100	0.6
3	107	1.2
Control	98	2.2

Although larger, the significance of variation between different sensors is lesser compared to single sample variations. The training technique used operates on the principle of change, so the nominal starting values are less significant. However, it still important to consider, and reduce where possible, the variability, to ensure operations are valid for all sensors. For example, if the digital filter used is designed for the bandwidth of a typical sensor pair, it must be considered whether this filter is also appropriate for any other credible variability in bandwidth.

Additionally, sensor adhesion method was not the focus of this project, and so with limited access to equipment and finite time, less practise and attention were focused on a methodology for repeatable attachment. In a real-world implementation, it is recommended that more attention should be given to the design of a robust methodology for sensor attachment, reducing variability between sensors.

4.3.3. Centre Frequency

The centre frequency was extracted from each of the frequency-domain plots. This was measured by selecting the frequency at which the maximum amplitude occurred.
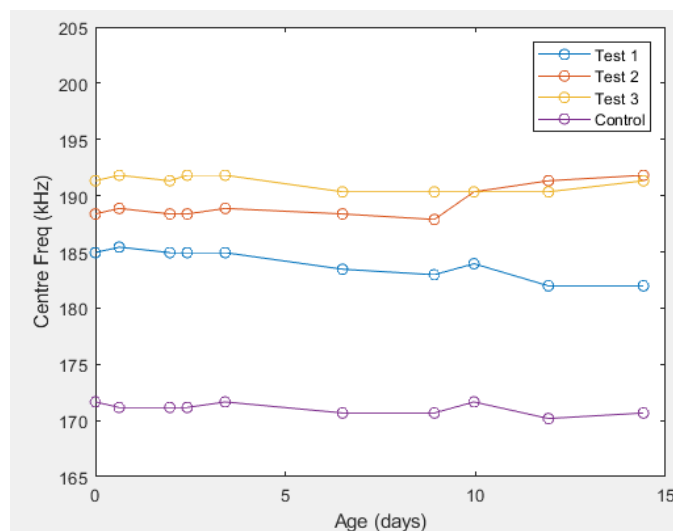

Figure 28 Centre frequency of first arrival signal as a funciton of age for tests one to three, corresponding to the three samples, as well as the control sample that was not exposed to temperature cycling

There was also no trend observed in the centre frequency for the three samples over the aging period (Figure 28). Although there is a large variability between samples of mean values, the standard deviation of recordings is comparable between the aged samples and control (Table 3).

Figure 38 to Figure 41 in the appendix show the spread of data points for individual time traces (before averaging) over the aging range.

Table 3 Mean and standard deviation of centre frequency for samples 1 to 3, as well as the control sample

Centre Frequency (kHz)		
Sample	Mean	Standard deviation
1	184	1.4
2	189	1.4
3	191	0.8
Control	171	1.5

4.3.4. Amplitude

Peak amplitude of the frequency-domain plot for each sample was extracted. This is known to be one of the signal features most affected by temperature.

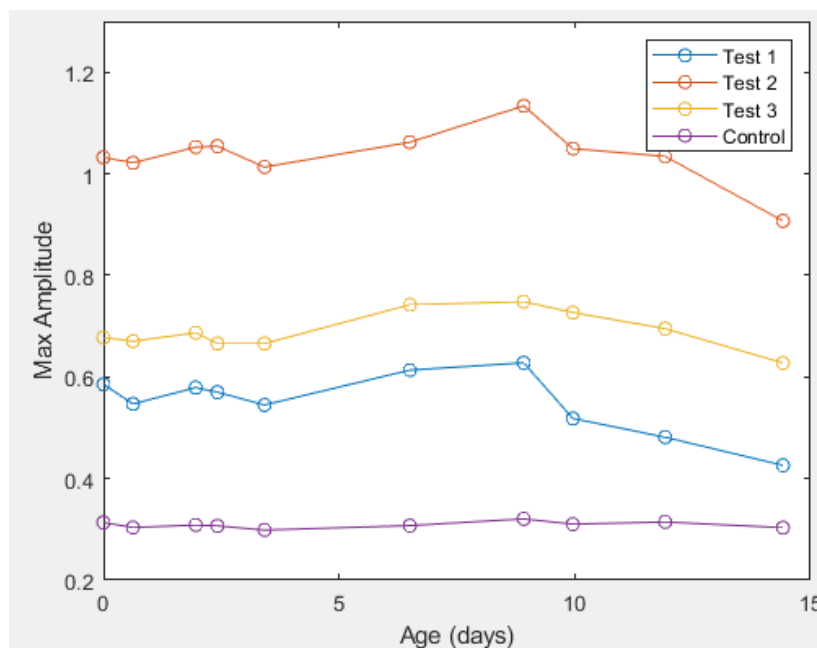*Figure 29 Max amplitude (with arbitrary units) of the frequency-domain plot, as a funciton of age for tests one to three, corresponding to the three samples, as well as the control sample that was not exposed to tmeperature cycling*

The results for the control sample, displayed in Figure 29, show a consistent peak amplitude across the testing period. For the aged samples, however, there is a slight trend of increasing

maximum amplitude from an age of 3 days up to 9 days, before then decreasing below the initial value. This was not observed in the control sample.

To evaluate this trend in further detail, the maximum amplitude metric was calculated for each of the 10 individual time traces per aging point before averaging.
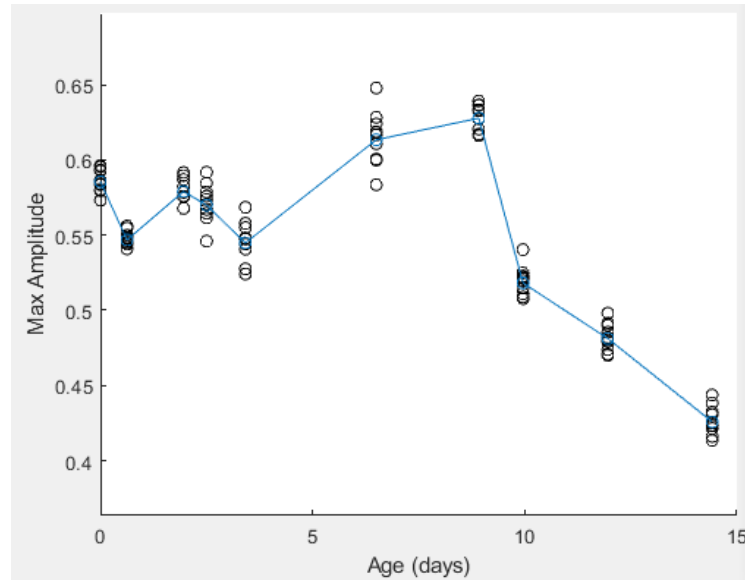
Figure 30 Maximum amplitude for sample 1, with blue circles representing measurements for each time-trace, and line representing the values for the averaged time trace.

Figure 30 shows a degree of variability of maximum amplitude around the value for the averaged time-trace. This is, however, is much smaller than the trend observed with heat treatment. This gives confidence in the significance of the observed change in signal amplitude as the sample, transducer and bond were progressively heat treated.

It was proposed that max amplitude of frequency-domain plot may be limited as a metric, as is it affected by bandwidth, with a broader spectrum having lower peak amplitude. Therefore, sum of squares (representing signal energy) was compared as a metric for overall amplitude.

Figure 31 Max amplitude (left) and sum energy of frequency domain plot (right) of the frequency-domain plot, as a function of age. Both plots have arbitrary y-axis units.

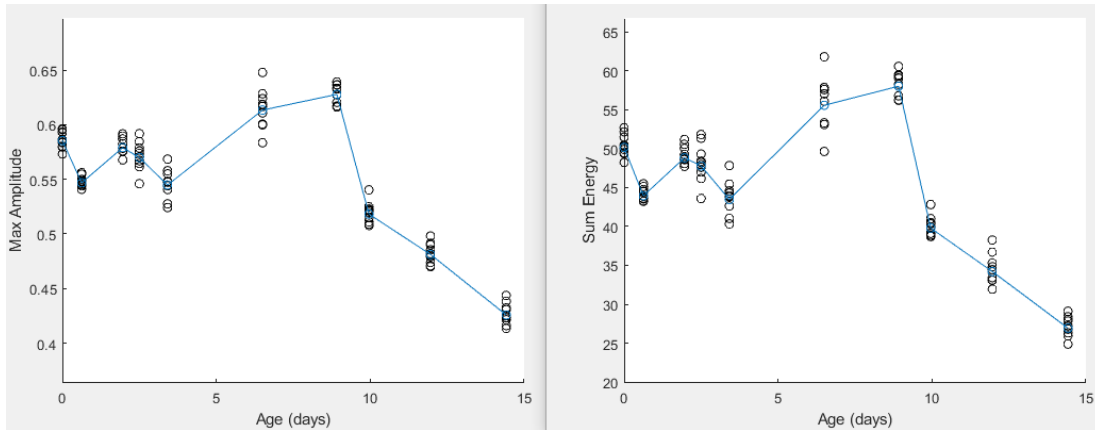Figure 31 shows a comparison of trends between the two metrics. It can be seen that the two metrics are comparable.

As amplitude is known to be affected by temperature during recording, it was important to establish whether the trends observed were due to an aging effect, or as a result of inconsistencies in cooling. If the sample and measurement system were not cooled to room temperature (the bond and transducer may still be above room temperature when the sample has cooled), then the aged samples would exhibit different amplitude values, not linked to aging effects.

4.3.5. Temperature when recorded

To evaluate recording temperature effects, where readings were not performed under identical conditions, a test was set up where the aged samples were removed from the high temperature environment before being recorded at various stages of cooling. Figure 32 shows the variation in the maximum amplitude of the frequency domain plot as a function of cooling time. It can be seen that the sample cooling is associated with a decrease in amplitude.
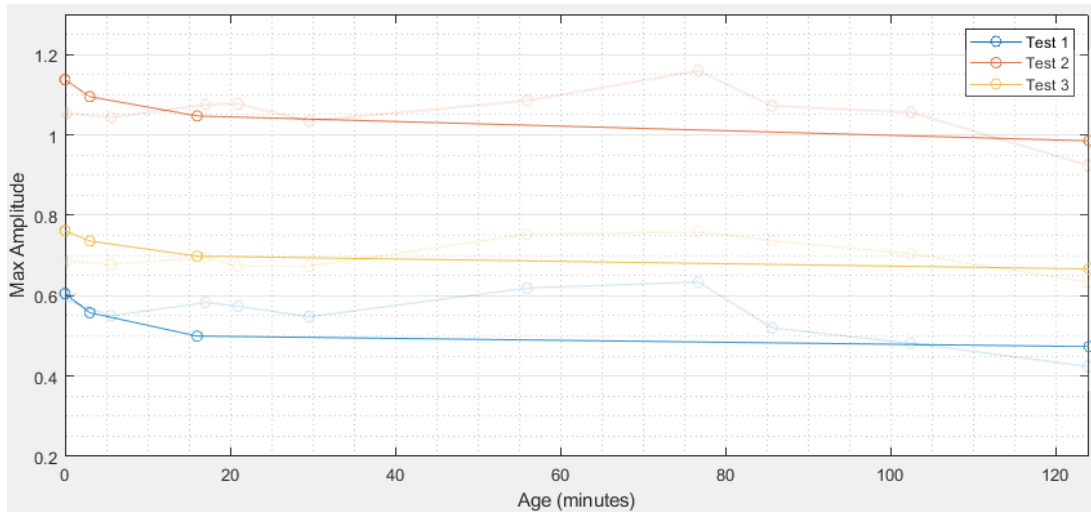
56

Figure 32 Max amplitude (with arbitrary units) as a function of cooling time (in minutes) for the three samples that were exposed to high temperature conditions. The aging results have been overlayed, with reduced opacity and x-axis scaling to permit comparison of magnitudes.

There was a 28% increase in amplitude at higher temperature than when fully cooled. The amount of cooling time for the first two data points was considerably less (and temperature of the sample and bond considerably higher) than during the aging trial. It can be, therefore, assumed that there was less variability of temperature in the aged sample tests, and so the amplitude effect on the results is less than 28%.

Additionally, as shown by the overlayed aging data in Figure 32, the range of amplitude values observed was still larger the range observed in the cooling trial, despite the cooling trial being recorded at a much greater range of temperatures. It can be concluded that some small component of the trend observed in the aging trial may be an effect of lack of control of temperature during recording, however there is another source of variability.

For completeness, centre frequency and bandwidth were also evaluated over the same cooling period.
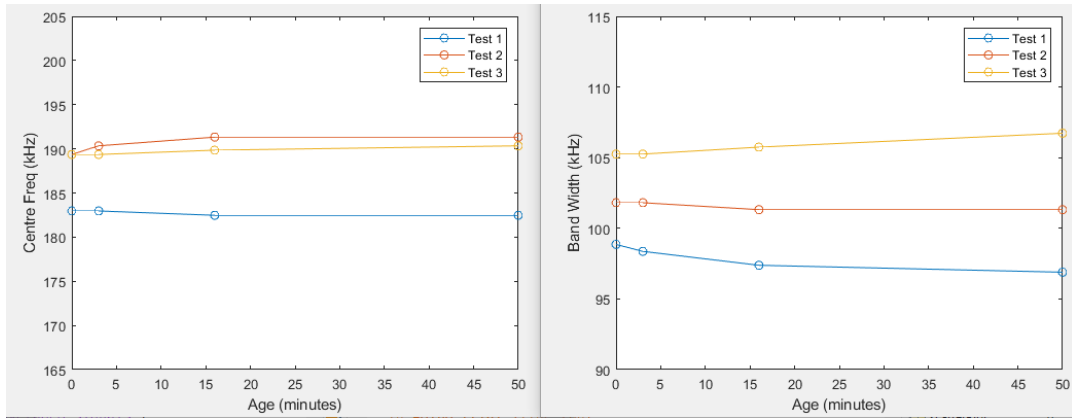
Figure 33 Centre frequency (left) and bandwidth (right) as a function of cooling time (in minutes) for the three samples that were exposed to high temperature conditions.

Figure 33 shows no significant trend between sample temperature and bandwidth or centre frequency. Additionally, different samples were affected differently by the temperature difference (sample 1 increasing in bandwidth, and sample 3 decreasing).

4.3.6. Assessment of longer-term effects

The final validation exercise in this work used to establish whether any trend exists, was to look at the metric values after a much longer aging period.

Following the two-week aging trial, the sample aging was continued (without regular data collection). A further data set was collected 70 days after sensor bonding.

Figure 34 Maximum amplitude (with arbitrary units) of the frequency-domain plot, as a function of age, including additional reading. Dotted lines indicate values at age = 0.

58

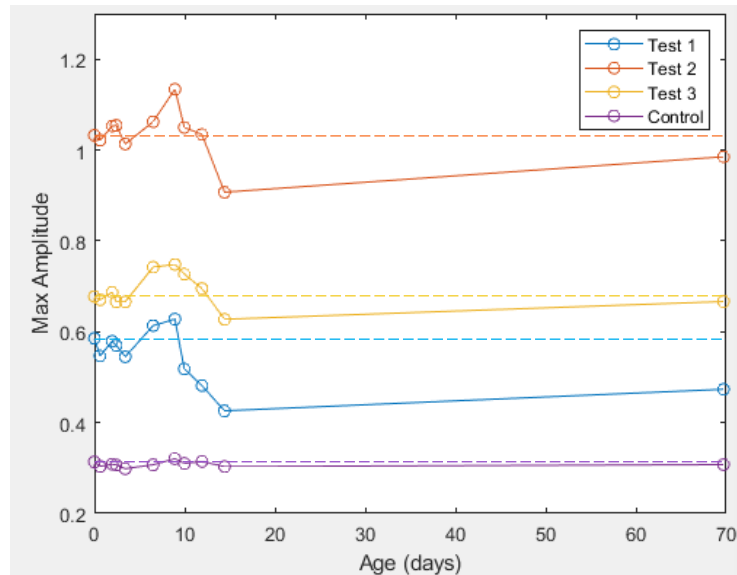Figure 34 shows the amplitude measure not continuing the trend observed in the earlier period, and instead increasing slightly toward the initial value (at age = 0). Therefore, it can be concluded that no lasting amplitude effect resulted from the aging period. It is clear, however, that some change in amplitude was observed that cannot credibly be attributed to temperature-at-recording (Figure 32). The control values varied by a very small degree (±6%) over the period of heat treatment, demonstrating that it was the heat-treatment that caused the significant variation in maximum amplitude across the aging period for all test (non-control) samples.

Bandwidth and centre frequency results are shown in Figure 35. These also show no discernible overall trend with age across the three samples.
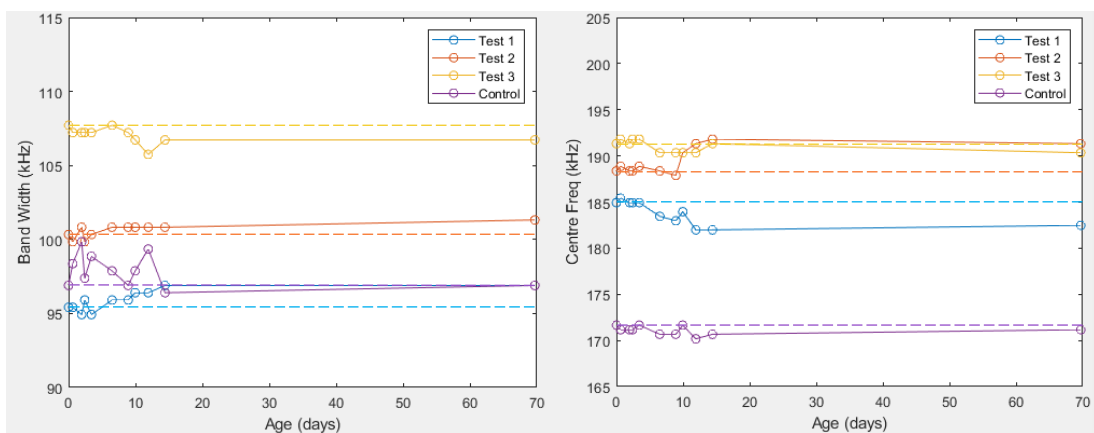
Figure 35 Bandwidth (left) and centre frequency (right) of first arrival signal as a funciton of age for three samples, including additional reading. Dotted lines indicate values at age = 0.

4.3.7. Baseline subtraction

The most important measure for this change in response, is the effect on the signal subtraction algorithm.

Figure 36 shows the heat-treated samples (1 to 3) exhibited a much greater change, as quantified by the maximum of residual after subtraction, compared with the control sample.

It should be noted that the first data point is omitted, as this reading was used as the reference baseline. Therefore, the first plotted data point (age = 2) represents the difference-metric change between a perfect match (reference at age = 1) and the recording after a first heat-treatment cycle.

Figure 36 Difference metric value compared to the first recorded signal (at age = 0) as a function of aging. Sample 1 (top left), sample 2 (top right), sample 3 (bottom left) and the control sample (bottom right). The black dotted line represents the worst match of the control sample (bottom right), as a reference. The orange line indicates the -25 dB threshold used in the training algorithm work (Section 3.6).
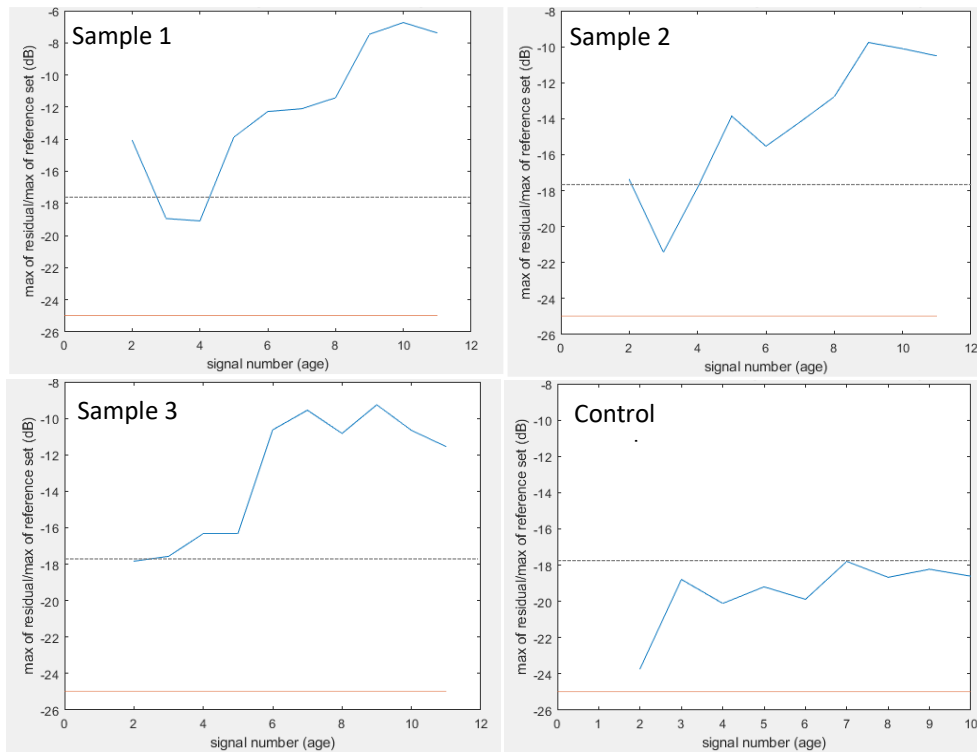
There is a change in the recorded signals over the heat-treatment period. This is observed in Figure 36, with later (heat treated) signals poorly matching with the signal at age = 0, or early in the heat treatment process. For samples 1-3, the greatest difference metrics between a signal recorded at age = 0, and subsequent signal after heat treatment, was between -6 to -10dB of the first arrival signal. This is compared to -18dB of the first arrival signal for the Control (for which the adhesive bond is still expected to have cured over the time-period used). It is noted that the manufacturer recommends a cure time of 16 hours at elevated temperature (40°C) for the bond to reach a stable strength (Araldite, 2011), however this cannot reasonably account for the substantial change of the heated samples relative to the control.

It is clear, then, that some other effect is present, other than the measures used in Sections 4.3.2, 4.3.3 and 4.3.4 (amplitude, bandwidth and centre frequency), since all these measures were shown to not exhibit a strong trend over the aging period. Additionally, the metric that most strongly correlated with age, maximum amplitude, will have been partially corrected for by the subtraction algorithm.
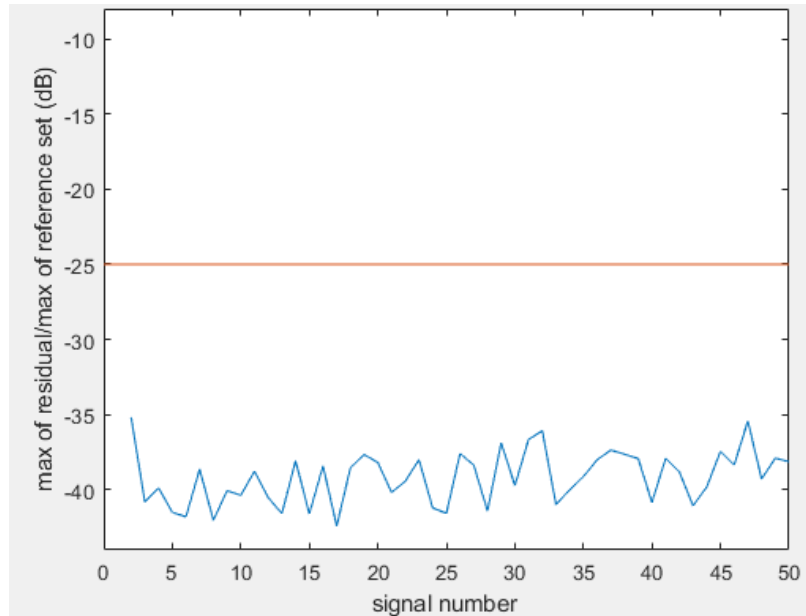
Figure 37 Difference metric value of 50 sequentially recorded signals, each created by averaging 10 signals, compared to a single baseline signal

Figure 37 displays the difference metric due to variations in equipment. The signals were all collected at the same age and under the same conditions. The difference between these signals is of the order of -40 dB, significantly lower than the effects observed in Figure 36.

4.4. Conclusions on the Accelerated Aging Trial

Three samples were exposed to high temperature cycling, with a fourth used as a control. A change was observed in the three heat-treated samples using the signal subtraction method presented in Section 3. This indicates that some accelerated aging was experienced by the sensors and bond, similar to that seen in the long-term duration data in Section 3.

In order to attempt to identify the specific signal change occurring with heat-treatment, three metrics were measured on a windowed section of the signal: bandwidth, centre frequency and amplitude.

Bandwidth and centre frequency were observed to be only slightly affected by temperature and the heat treatment process. The amplitude of the windowed section of the signal did, however, vary significantly over the course of the heat treatment. The observed change was an increase in amplitude across all three aged samples, before reducing back to near the original value before heat treatment.

It was hypothesised that this trend may have been a result of differing sample temperatures at recording, since transducer amplitude behaviour is known to vary with temperature. However, an investigation into response as a function of cooling time (and so recording

temperature) demonstrated that the temperature-dependent effect was not sufficient to wholly explain the trend observed in heat treatment. The data suggests, then, that heat treatment affects amplitude, however, an additional aging effect occurs that is not represented by bandwidth, centre frequency or amplitude, since all these measures returned to their original value (before heat-treatment).

The reduction in signal subtraction performance with heat treatment, as described above, may be similar to, and somewhat representative of, the change observed in the long-duration data previously collected, however, there is insufficient information to be able to determine this.

5. Overall Conclusions and Recommendations for Further Work

5.1. Conclusions

There are many naval applications for Structural Health Monitoring solutions, and despite many recent advances in the enabling technologies, quick adoption into industry still appears unlikely. Many technical challenges and areas of insufficient understanding still exist that present risk when businesses consider significant investment into these systems. Qualification, and long-term operation are two significant obstacles.

It is likely that qualification will require a blend of NDE qualification methodologies, where each parameter and part of the system is understood, and data-driven technology qualification, where statistical analysis of results may be used to prove technologies that cannot be decomposed easily into demonstrable sub-tasks (often referred to as "black box methods"). The first part of this project concerned the understanding of the behaviour of a data processing algorithm on some previously recorded long-term data collected on a structure.

Many Structural Health Monitoring methodologies require an initial training period to model the 'undamaged state' of the asset, from which deviations can be measured, and flagged as potential damage.

The model of the undamaged state in this work is described by a set of baseline guided-wave ultrasonic testing signals recorded on an undamaged structure. The baseline stretch aspect of this method is used to fill gaps in recorded baselines, and to reduce the number of baselines required to model the undamaged state. The stretch involves a number of operations on recorded signals to emulate the effects of varying environmental and operational conditions.

The greater the number of shift operation combinations trialled, the more likely a close match will be made between a baseline and new signal. Computational savings can be made, however, by assessing the most commonly used parameters for each operation. Additionally, this restricts the behaviour of the system so that each part of the model operates as predicted. That is, that each model is designed to model and compensate for a particular behaviour, so this ensures there is no overlap in role of each operation that may break the limits of applicability of each model.

Depending on the application and availability of data, a balance of performance and computational burden/understood behaviour can be achieved through the control of interaction and extent of each shifting operation. In all cases, however, a larger number of

training datasets (resulting in larger coverage of baselines to model the normal state) was proposed to provide a much more robust system, permitting any baseline shifting operations to be used to accommodate for only small changes in conditions.

One of the most significant limiting factors on the performance of the developed algorithm, and all methodologies based around making an initial baseline model of an asset state, is the benign change in sensors and their bonds that cannot be differentiated from defect growth. Therefore, the understanding and demonstration of SHM system performance over the timeframe of a realistic vessel service life is crucial to their transition from research to industry use. The second part of this project looked to further the understanding of aging effects that are present in long-term duration monitoring, and how to replicate them in practical time scales.

A trial was undertaken to attempt to mimic the long-term changes on sensor systems in a shorter period of time. It was hoped that accelerated aging may be induced through the use of high temperature heat-treatment. A number of plate samples had transducers bonded to them, and were heat-treated, with measurements made at intervals to observe the effect on the recorded data from any heat-treatment induced changes.

The samples that were heat treated did show a marked change over the heat-treatment period when compared to the control sample. Using a signal-subtraction-based measure of signal similarity, recordings made at the beginning and end were found to be poorly matched. The behaviour observed in the heat-treatment trial may have been a similar aging effect was to that observed in the long-duration recording. The specific data change that results in the reduced subtraction performance, however, is still yet to be determined, since this work found that amplitude, bandwidth and centre frequency were not reliable measures of the extent of the heat treatment on the samples.

In conducting this work, many valuable results and methodologies were identified that further the understanding of the behaviour of Structural Health Monitoring systems. Much work remains, however, for industry-ready systems to be suitable for the replacement of current processes use to assess naval structures.

5.2. Further work

Perhaps the most valuable activities that could be conducted in the area of guided wave SHM is the conducting of long-term experiments or trials under controlled conditions. The availability of many different data sources monitoring as much about the potential effectors on system behaviour, means the performance, limitations and behaviour of SHM systems

can be understood and verified. This includes the independent measurement of the structural health with other methods (as well as the SHM system) during the trial, to permit conclusions to be drawn on sources of change observed. Additionally, observations about structure change (for example noting of areas of localised corrosion or loss of coatings), as well as recording of environmental and operating data to cross-reference to changes observed SHM data.

For understanding of expected performance in a real application, the use of realistic data (and so trials) is vital. Without this it is difficult to quantitatively assess the viability of these methods for the replacement and reduction of current NDE practises.

This work did not use signals containing damage, either physical or synthetically added. To properly assess a guided wave SHM system, the damage type and variability must be included in trials. Using principles of guided wave analysis, knowledge of component manufacture and typical causes of failure, as well as structural integrity assessment, application-specific criteria for indication sentencing may be determined. This can then be used to evaluate system performance against.

The damage metrics considered for this work were limited to the basic measures presented in previous literature, RMS and maximum of a residual signal after subtraction. Both of these metrics have a different suitability, depending on the target reflector and complexity of the signal. For many real-world applications, it is necessary to detect different defect types, or both techniques may be of only limited suitability. Therefore, there is believed to be merit in the investigation into the use of combination difference-metrics, using both RMS and maximum amplitude of residual signals. Furthermore, the use of relative differences in metrics from each flaw type and morphology, may build an understanding that could increase ability for reliable characterisation and more complex sentencing rules.

Due to the constraints on collection of data, and since synthetic damage was not considered for this work, there is a lack of the statistical analysis that would be required to evaluate probability of detection. Statistical analysis and modelling constitute a large part of the demonstration and evaluation of performance, and there is considerable value in this as future work.

Certain improvements are proposed for the hardware system, which were not possible due to the time and equipment constraints on this work. The method of adhesion was not ideal in the trial, and the repeatability and consistency of bond would benefit from improvements in cleanliness and clamping pressure. Performing the bonding process in a vacuum bag is proposed as a method of degassing the bond to reduce contamination, and the known

pressure on the transducers during adhesive setting is expected to provide consistent bond properties.

The electronics design was likely to introduce some noise from impedance mismatches. This was not expected to have a significant effect on this investigation due to its aims of measuring change over an aging period. However, for the design of a monitoring system, more care would need to be given to the design of the electronic circuit.

Another area of work that was not investigated as part of this project, but that is vital for the implementation of a practical monitoring system, is the use of an array of sensors. This both enables larger area coverage but permits significantly more information to be elicited about the structure in the collection of the "full matrix" of transmitter and receiver pairs from an array. The result of this is the improved localisation and characterisation of indications where detection occurs in multiple recordings.

6. References

Alleyne, D. & Cawley, P., 1992. Optimization of lamb wave inspection techniques. *NDT&E International,* Volume 25, pp. 11-22.

Araldite, 2011. *Araldite Rapid Two component fast cure epoxy adhesive Datasheet.* [Online]
Available at: https://static.rapidonline.com/pdf/87-4038_v1.pdf
[Accessed 1 April 2022].

Araldite, 2014. *ARALDITE STANDARD 2 X 15ML.* [Online]
Available at: https://www.go-araldite.com/products/epoxy-adhesives/araldite-standard-2-x-15ml-tube
[Accessed 16 August 2021].

Attarian, V. A., Cegla, F. B. & Cawley, P., 2014. Long-term stability of guided wave structural health monitoring using distributed adhesively bonded piezoelectric transducers. *Structural Health Monitoring,* 13(3), pp. 265-280.

Cawley, P., 2003. *Practical long range guided wave inspection – applications to pipes and rail.* Channai, NDE2002 National Seminar or INST, pp. 66-74.

Cawley, P., 2018. Structural health monitoring: Closing the gap between research and industrial deployment. *Structural Health Monitoring,* 17(5), pp. 1225-1244.

Cawley, P., Cegla, F. B. & Galvagni, A., 2012. Guided waves for NDT and permanently installed monitoring.. *Insight - Non-Destructive Testing and Condition Monitoring,* 54(11).

Clarke, T., Simonetti, F. & Cawley, P., 2010. Guided wave health monitoring of complex structures by sparse array systems: Influence of temperature changes on performance. *Journal of Sound and Vibration,* 329(12), pp. 2306-2322.

Courtier, 2018. *EngD - Guided Wave Structural Health Monitoring of Complex Aerospace Components.* Bristol: The University of Bristol.

Croxford, A. J., Moll, J., Wilcox, P. D. & Michaels, J. E., 2010. Efficient temperature compensation strategies for guided wave structural health monitoring. *Ultrasonics,* 50(4-5), pp. 517-528.

Croxford, A. J., Wilcox, P. D., Drinkwater, B. W. & Konstantinidis, G., 2007. Strategies for guided-wave structural health monitoring. *Proceedings of the Royal Society A Mathematical Physical and Engineering Sciences,* 463(2087), pp. 2961-2981.

Friedmann, H. & Kraemer, P., 2016. *Vibration-based condition monitoring, structural health monitoring, population monitoring – Approach to a definition of the different concepts by means of practical examples from the field of wind energy.* Bilbao, Spain, 8th European Workshop On Structural Health Monitoring (EWSHM 2016).

Gorilla Glue, 2019. *GORILLA EPOXY.* [Online]
Available at: https://uk.gorillaglue.com/gorilla-epoxy/
[Accessed 16 August 2021].

Harley, J. & Moura, J. M. F., 2012. Scale transform signal processing for optimal ultrasonic temperature compensation. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control,* 59(10), pp. 2226-2236.

Herdovics, B. & Cegla, F., 2018. Long-term stability of guided wave electromagnetic acoustic transducer systems. *Structural Health Monitoring,* 19(1), pp. 3-11.

Krautkrämer, J. & Krautkrämer, H., 1990. *Ultrasonic Testing of Materials.* 4th ed. Berlin: Springer.

Lu, Y. & Michaels, J. E., 2005. A methodology for structural health monitoring with diffuse ultrasonic waves in the presence of temperature variations. *Ultrasonics,* 43(9), pp. 717-731.

Permabond, 2014. *Heat-Resistant, Two-Part Epoxy.* [Online]
Available at: https://www.permabond.com/resource-center/high-temperature-resistant-2part-epoxy/
[Accessed 16 August 2021].

Putkis, O., 2014. *Ultrasonic GWSHM and its application to anisotropic composite material.* Bristol: The University of Bristol.

SAE International, 2013. *Guidelines for implementation of structural health monitoring, ARP6461.* Warrendale(Pennsylvania): s.n.

Zhao, Y. et al., 2017. Generation mechanism of nonlinear ultrasonic Lamb waves in thin plates with randomly distributed micro-cracks. *Ultrasonics,* Volume 79, pp. 60-67.

7. Acronyms

RF	Radio Frequency
SHM	Structural Health Monitoring.
UT	Ultrasonic Testing
PZT	Lead Zirconate Titanate
NDT	Non-Destructive Testing
NDE	Non-Destructive Evaluation
SNR	Signal to Noise Ratio
BSS	Baseline Stretch Subtraction
OBS	Optimum Baseline Subtraction
DC	Direct Current
FFT	Fast Fourier Transform

8. Appendix
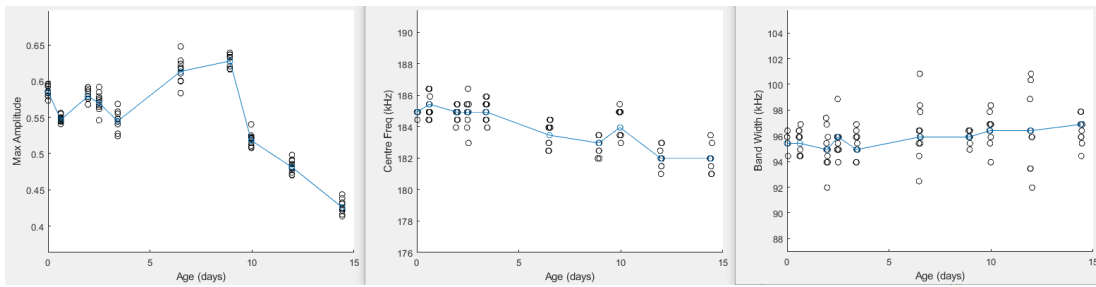
Figure 38 Sample 1: Max amplitude of frequency plot for first arrival signal (left), frequency of max amplitude (centre), and -6 dB bandwidth about the centre frequency (right)
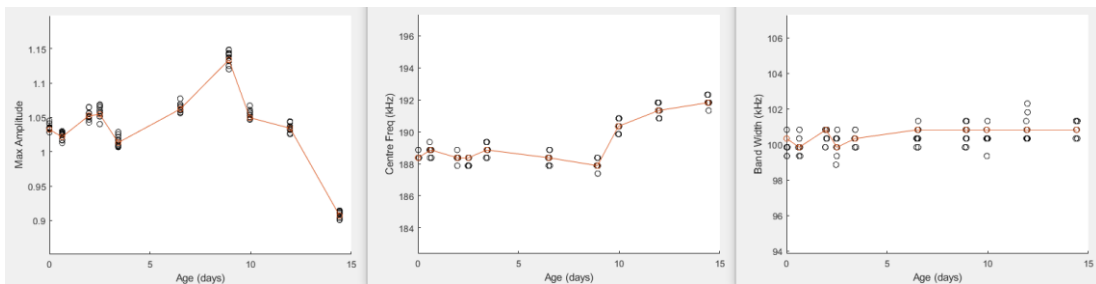

Figure 39 Sample 2: Max amplitude of frequency plot for first arrival signal (left), frequency of max amplitude (centre), and -6 dB bandwidth about the centre frequency (right)
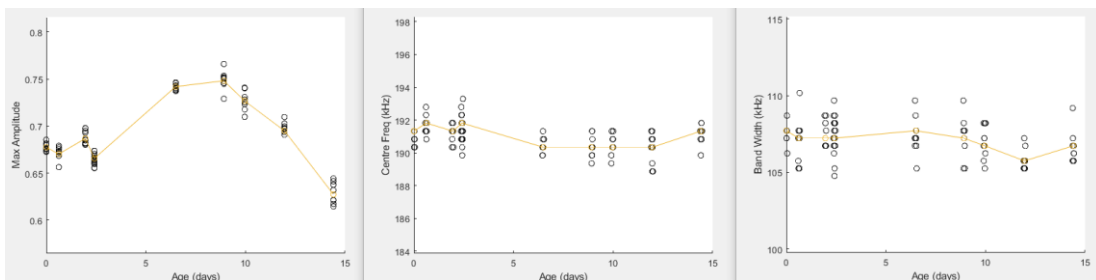

Figure 40 Sample 3: Max amplitude of frequency plot for first arrival signal (left), frequency of max amplitude (centre), and -6 dB bandwidth about the centre frequency (right)
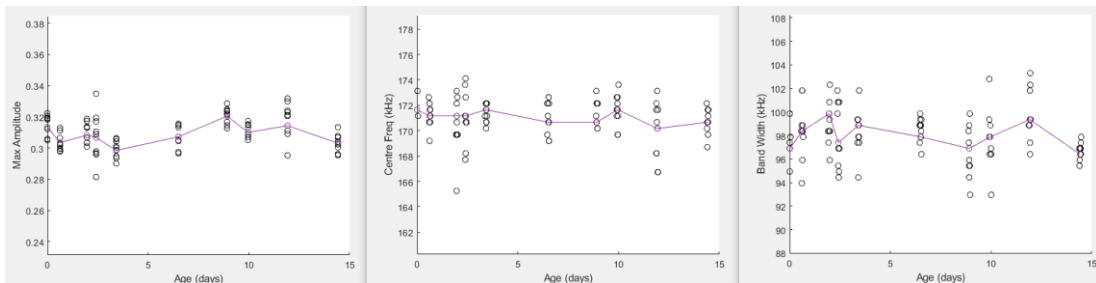

Figure 41 Control sample: Max amplitude of frequency plot for first arrival signal (left), frequency of max amplitude (centre), and -6 dB bandwidth about the centre frequency (right)