Fusion of multi-view ultrasonic data for increased detection performance in non-destructive evaluation

Paul D. Wilcox[1], Anthony J. Croxford[1], Nicolas Budyn[1], Rhodri L. T. Bevan[1], Jie Zhang[1], Artem Kashubin[2], Peter Cawley[2]

[1]Department of Mechanical Engineering, University of Bristol, Queen's Building, University Walk, Bristol, BS8 1TR, UK.

[2]Department of Mechanical Engineering, Imperial College, Exhibition Road, London, SW7 2AZ, UK.

ABSTRACT

State-of-the-art ultrasonic Non-Destructive Evaluation (NDE) uses an array to rapidly generate multiple, information-rich views at each test position on a safety-critical component. However, the information for detecting potential defects is dispersed across views, and a typical inspection may involve thousands of test positions. Interpretation requires painstaking analysis by a skilled operator. In this paper various methods for fusing multi-view data are developed. Compared to any one single view, all methods are shown to yield significant performance gains, which may be related to the general and edge cases for NDE. In the general case, a defect is clearly detectable in at least one individual view, but the view(s) depends on the defect location and orientation. Here, the performance gain from data fusion is mainly due to the selective use of information from the most appropriate view(s) and fusion provides a means to substantially reduce operator burden. The edge cases are defects that cannot be reliably detected in any one individual view without false alarms. Here certain fusion methods are shown to enable detection with reduced false alarms. In this context, fusion allows NDE capability to be extended with potential implications for the design and operation of engineering assets.

1 INTRODUCTION

Multi-element ultrasonic arrays are widely used for the Non-Destructive Evaluation (NDE) of safety-critical engineering components and structures across a range of industries [1]. Arrays were first used in NDE to replicate inspections that had previously been performed by scanning single-element, monolithic transducers. To achieve this, physical beamforming is performed by exciting multiple elements in an array with delays chosen so that the emitted ultrasonic beam mimics that from a monolithic transducer. In this mode of operation, arrays enable one or more mechanical scanning directions to be replaced by electronic scanning of the active aperture in the array. Furthermore, the same array can be used to focus ultrasound at different depths, thus enabling inspections that require separate monolithic transducers with different focal depths to be performed with one array.

Increasingly, ultrasonic arrays are used to perform Full Matrix Capture (FMC) acquisition where each element is fired in turn while the ultrasonic A-scans are recorded on all elements [2]. In FMC mode, there is no physical beamforming at acquisition and instead all beamforming is performed in post-processing. This enables imaging algorithms to be implemented that are either infeasible or impossible to implement through physical beamforming. These include the delay-and-sum Total Focusing Method (TFM) [2] and frequency-domain variations such as inverse wavefield extrapolation (IWEX) [3] and the wavenumber algorithm [4]. In such algorithms the entire aperture of an array is focused in transmission and reception at every imaging point. For this reason, they can attain the theoretical diffraction limit for linear imaging resolution. Also relevant to the technique proposed in the current paper are Plane Wave Imaging (PWI) methods where a reduced number of transmission cycles are used to emit plane waves at various angles with focusing being performed in reception for each emission and the results superposed. PWI yields images of comparable quality to those obtained from FMC data but with substantially fewer transmission cycles. Again PWI techniques can be implemented as time-domain delay-and-sum formulations or frequency-domain migrations [5].

The current paper is concerned with multi-view imaging, where multiple images of the same physical region in space are formed from one set of data by exploiting different mode-conversions and reflections. Multi-view images can be formed by any suitable imaging algorithm (e.g. TFM [6] or PWI imaging [7]). The salient point is that the sensitivity to a defect depends on the view as well as the position, type and orientation of the defect. The concept of using information from multi-view images has been considered previously. Han *et al.* [8] simulated direct, half-skip and full-skip longitudinal mode TFM images of a non-planar branched crack. From the separate direct and full-skip images (the half-skip image was discarded), the positions and widths of any peaks

were identified and plotted on a single graph. This gave a reasonable qualitative reconstruction of the crack shape. Horchens *et al.* [9] used a pair of oblique incidence ultrasonic arrays to inspect a weld from both sides using mode-converted shear waves in the component. By exploiting reflections of the top and bottom surfaces of the component and using the pair of transducers both as a pitch-catch pair and as individual pulse-echo devices, 10 different views of the weld were obtained using IWEX imaging. The different views were coded by colour and superposed as a single composite image. It was qualitatively observed from the colour-coded composite image that different views were sensitive to different aspects of the different defects. For example, some enabled the specular reflection from the flank of an artificial crack-like defect to be observed directly, while others only showed tip diffraction signals. The possibility for mis-registration between views due to uncertainty in the exact component geometry was noted. Sy *et al.* [10] proposed the use of a Specular Echoes Estimator (SEE) function, which provides an estimate of the amplitude of response that would be obtained at each point in each view if a planar specular reflector in a specified orientation existed at that point. The SEE functions for different views are then examined to determine the most appropriate view to image defects of the specified orientation in any region. Noise is not explicitly measured, but the implicit assumption is that noise is constant in all views, hence the SEE function is a proxy for Signal to Noise Ratio (SNR).

It is clear from [6]-[10] that multi-view imaging provides additional information beyond that obtained in a single view that may aid manual characterisation of defects. It also seems evident that the information in multi-view images should enable improved detection performance. In most NDE applications, defects occur infrequently and hence manual analysis of multi-view images is a viable strategy for defect characterisation once a defect is detected. However, reliable manual analysis of multi-view images to improve detection is likely to be impractical and unreliable due to the volume and complexity of data that must be analysed. For example, a circumferential scan of a butt weld in a 12” diameter pipe at a 1 mm scan pitch will produce 1000 datasets. If direct, half-skip and full-skip paths are considered, 21 multi-view images can be generated from each dataset so an operator must potentially view 21,000 images per weld. The current paper concerns the specific question of how the information in multi-view images may be automatically combined through data fusion to improve detection performance. Defect characterisation is not considered. The objective of detection is to determine if there is an indication of one or more defects in the multi-view images generated from a dataset. If an indication is found, then appropriate methods for the characterisation of the defect(s) needs to be deployed.

Ultimately, the motivation for applying data fusion is always to improve overall inspection performance, but it is instructive to consider this in terms of general and edge cases. The general case is the detection of defects that are readily visible if the correct view is selected, but where the correct view depends on the position, type and orientation of the defect; no one single view enables the reliable detection of all defects of interest. The desired outcome of data fusion in the general case is to ensure that defects manifesting in any view are detected, but crucially without increasing the probability of false alarm. The edge case for multi-view data fusion is when a defect is barely visible in any individual view. Here the desired outcome is to increase the reliability of detecting such a defect (i.e. reducing the probability of false alarm for a given probability of detection).

The paper is arranged as follows. Section 2 describes the inspection configuration and the information required for data fusion, including the generation of the multi-view images. In Section 3 a common mathematical framework is developed for various candidate data fusion methodologies. The fusion methods are based on expected statistical distributions of noise and, in some cases, defect responses. The defect responses are obtained using an approximate defect scattering model and various simplifying assumptions. In Section 3, the performance of the data fusion methods compared at a single image point using numerical data drawn from the expected statistical distributions, and hence has exactly the statistical properties expected. In Section 4 the candidate techniques are applied to image data that is independently synthesised from experimental measurements and finite element simulations; here the data has statistical properties that are more representative of those likely to be encountered experimentally, and do not necessarily match the expected distribution. Sections 3 and 4 show the potential gains of using data fusion in both the general and edge cases.

In Section 5, the candidate methods are applied to experimental data obtained from an array being scanned past a small defect in a highly-scattering material, giving a practical demonstration of the benefit of data fusion in an edge case.

2 EXPERIMENTAL FRAMEWORK AND CREATION OF DATA FOR FUSION

2.1 Inspection configuration

To demonstrate the data fusion principle, an oblique incidence configuration shown in Fig. 1(a) is selected. A commercial 128-element linear array (Imasonic, Voray-sur-l'Ognon, France) with 0.63 mm element pitch and a nominal centre frequency of 5 MHz is used. Data is collected in Full Matrix Capture (FMC) mode using an array controller (Micropulse, Peak NDT Ltd., Derby, UK) at a sampling frequency of 25 MHz. The raw FMC data is transferred to a computer where all processing is performed. Prior to imaging, the FMC data is digitally filtered (using a 5 MHz Gaussian filter with -40 dB points at 0.5 and 9.5 MHz) and converted to an analytic signal using a Hilbert transform to generate the imaginary quadrature component. A detailed description of the signal processing of the raw measured data performed prior to imaging is provided in Section S.1 the supplementary material.

In the current paper, water is used as the coupling medium, although the same approach is applicable to coupling though a solid wedge. The oblique configuration is commonly used in industry for the inspection of critical welds. In such a scenario, the Region Of Interest (ROI) contains the cross-section of the weld with the weld running perpendicular to the plane of the diagram. The array is typically scanned along the weld to form cross-sectional images of the weld at each scan position. The defects of interest in a weld include: fusion-face cracks at the weld-parent metal interface; porosity; inter-run defects in the body of the weld; and concavity or protrusion at the weld root.

The oblique configuration is ideal for generating multi-view images as mode conversions at the front and back walls of the component (i.e. top and bottom surfaces in Fig. 1(a)) can be exploited. Here, the first 21 unique views are considered; those which contain duplicate data due to acoustic reciprocity are ignored (e.g. L-T and T-L are duplicates so only the former is required). The 21 views considered include at most one reflection off the bottom surface of the component on either the transmission or reception path, or both. This is illustrated in Fig. 1(b), which also describes the naming convention used to describe the different views. Accurate imaging and accurate co-registration of multi-view images requires knowledge of all inspection parameters, including array position relative to the specimen, the specimen thickness and all ultrasonic velocities. For the purposes of the current paper, these were independently measured to high accuracy using appropriate methods.

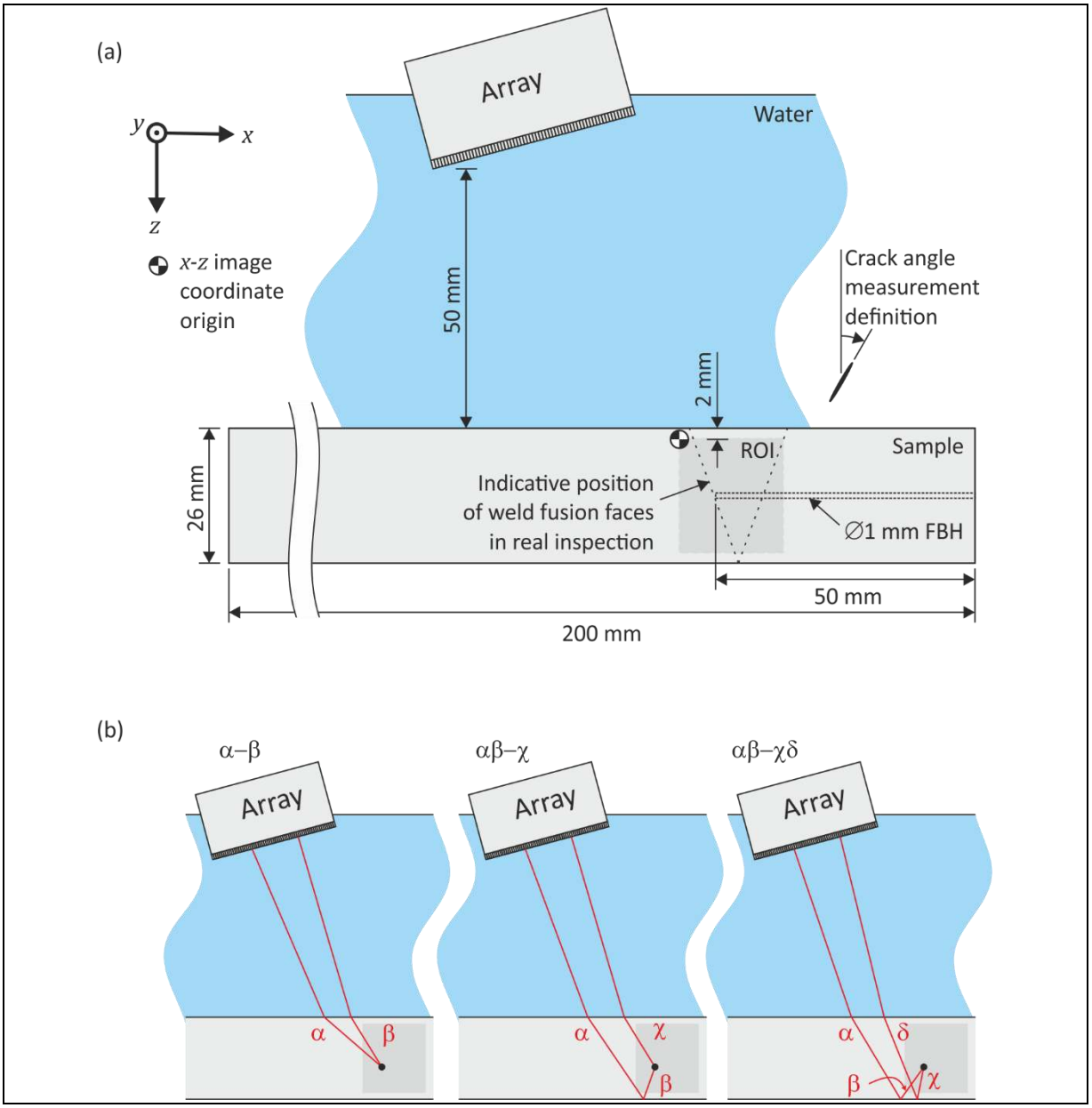

Fig. 1 (a) Inspection configuration showing Region Of Interest (ROI) and location of Flat Bottom Hole (FBH) in experimental sample; (b) multi-view naming conventions, where the Greek letters can be either L or T representing longitudinal and shear wave modes respectively.

2.2 Creation of multi-view images

The process of forming multi-view TFM images is described in detail elsewhere [6]. The image value at position $\mathbf{r}$ in i^{th} view is given by:

$$I_i(\mathbf{r}) = \sum_{T,R} A_{TR}^{(i)}(\mathbf{r}) h_{TR}\left(\tau_{TR}^{(i)}(\mathbf{r})\right) \tag{1}$$

where $A_{TR}^{(i)}(\mathbf{r})$ and $\tau_{TR}^{(i)}(\mathbf{r})$ are the amplitude coefficients and time delays for the view, $h_{TR}(t)$ are the filtered analytic FMC signals, and subscripts T and R denote transmitter and receiver elements. Here, no apodisation is used and $A_{TR}^{(i)}(\mathbf{r}) = 1$. For the purposes of the work described in the current paper, it is assumed that any mis-registration between views is small compared to the size of the imaging Point Spread Functions (PSFs). At typical ultrasonic NDE frequencies, this corresponds to millimetre-order co-registration accuracy. A strategy for accommodating mis-registration is discussed in Section 6.1. Note that although the numerical image values,

$I_i(\mathbf{r})$, are complex, the phase information is discarded and instead only the amplitudes $|I_i(\mathbf{r})|$ are considered for data fusion. This is because in practice it is difficult to achieve co-registration at the desired level of accuracy for phase between views at a point to be reliably exploited. The images in the left half of Fig. 2 show example multi-view data, $|I_i(\mathbf{r})|$, obtained from the experimental configuration shown in Fig. 1(a). These results are presented on a common 40 decibel (dB) scale relative to the largest signal present in any view.

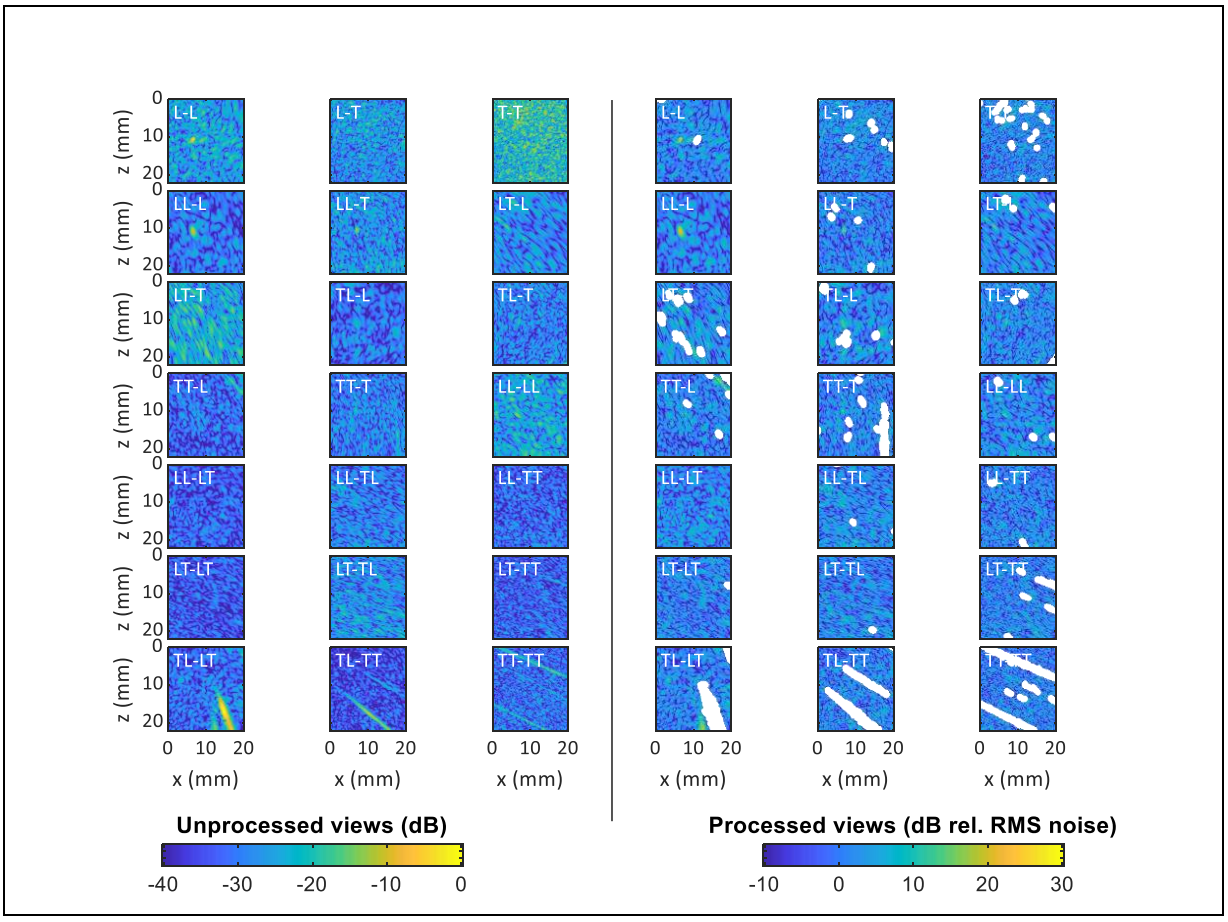

Fig. 2 Example multi-view results obtained from experimental measurements on a copper sample in a region containing the ∅1 mm flat-bottomed hole (the tip of the hole is intended to represent a small circular crack in the $y-z$ plane) shown in Fig. 1(a). The images on the left-hand side of the figure are the unprocessed results of multi-view imaging expressed on a common dB scale relative to the largest response present in any view. The images on the right-hand side of the figure are the same results after multiplying by $c_i(\mathbf{r})$ (correction function for beam spreading and attenuation) and $m_i(\mathbf{r})$ (mask function to remove artefacts). These images are expressed on a dB scale relative to the RMS noise level ($\sqrt{2}\sigma_i$) in each view, with masked regions indicated in white. The mask and correction functions for each view are automatically calculated from a dataset obtained in a pristine region.

2.3 Creation of input data for fusion

A naïve method of fusing data from multi-view images, such as simple summation, i.e. $\sum_i |I_i(\mathbf{r})|$, takes no account of (a) the noise level, (b) the expected defect response, or (c) the presence of imaging artefacts due to the inspection geometry in different views. Consequently, the fused result is not optimal, and may provide worse defect detection performance than the individual contributions.

The process proposed in the current paper involves using one or more reference datasets obtained on a pristine region of the component, from which a set of reference multi-view images can be generated. These images contain coherent speckle noise (due to material microstructure) and imaging artefacts. The latter are due to reflections from structural features along one ray path (e.g. the upper and lower surfaces of the specimen) appearing in a view associated with a different ray path at non-physical locations. Examples of such artefacts are clearly visible in the TL-LT, TL-TT and TT-TT views on the left half of Fig. 2. Artefacts are removed prior to data

fusion by masking the regions in each view where artefacts are found to occur in a dataset obtained from a pristine specimen. The mask function for the i^{th} view, $m_i(\mathbf{r})$, is defined as 0 in artefact regions and 1 in the artefact-free regions of interest. The hypotheses behind the automated process to remove artefacts are

(a) that in a pristine specimen, the image in the $m_i(\mathbf{r}) = 1$ region in the i^{th} view contains only microstructural noise that manifests as image speckle, described at each point by a Rayleigh distribution [11] with Rayleigh parameter $\sigma_i(\mathbf{r})$;
(b) that the spatial variation in $\sigma_i(\mathbf{r})$ due to beam-spreading and attenuation in a pristine specimen can be approximated as a log-linear function of position, i.e. $\ln \sigma_i(\boldsymbol{r}) \cong \ln \sigma_i + \mathbf{b}_i \cdot \mathbf{r}$, where σ_i and $\mathbf{b}_i$ are constants for the view;
(c) that multiplication of each view by $\exp(-\mathbf{b}_i \cdot \mathbf{r})$, therefore results in speckle in the $m_i(\mathbf{r}) = 1$ region with a spatially-uniform Rayleigh-distributed amplitude described by Rayleigh parameter σ_i.

The method is described in detail in [12], but the basic principle (applied separately to each view) is the iterative growth of the $m_i(\mathbf{r}) = 0$ region until the image in the $m_i(\mathbf{r}) = 1$ region satisfies the above hypotheses. At each iteration $\mathbf{b}_i$ and σ_i are recalculated to fit to the image amplitude in the $m_i(\mathbf{r}) = 1$ region; the image in this region is then multiplied by $\exp(-\mathbf{b}_i \cdot \mathbf{r})$ and the amplitude distribution compared to a Rayleigh distribution. If the match is poor, the $m_i(\mathbf{r}) = 0$ region is expanded to include the next highest amplitude points and the next iteration starts. Once a satisfactory match to a Rayleigh distribution is obtained the process yields the mask function, $m_i(\mathbf{r})$, the amplitude correction function, $c_i(\mathbf{r}) = \exp(-\mathbf{b}_i \cdot \mathbf{r})$, and an estimate of the Rayleigh parameter for the speckle noise, σ_i in each view. The accuracy of the latter estimate is governed by the number of independent values that remain in the $m_i(\mathbf{r}) = 1$ region; typically this is several hundred (based on the size of the ROI and assuming a speckle correlation length equal to the wavelength) and the error in the estimate is therefore 5% or less. For the purposes of subsequent data fusion, it is assumed that this error is negligible compared to other sources of uncertainty and σ_i is treated as a known deterministic quantity.

For subsequent datasets the masked and corrected response in each view is

$$x_i(\mathbf{r}) = c_i(\mathbf{r}) m_i(\mathbf{r}) |I_i(\mathbf{r})| \tag{2}$$

and this forms the main input to the all the data fusion processes described in the next section. The results of applying this correction and masking to multi-view data are shown in the right half of Fig. 2. The corrected and masked multi-view images are again presented on a 40 dB scale, but this time, 0 dB is the RMS noise level for each view. The consequence of applying the correction is that the RMS noise level should be spatially constant in each view and this is clearly demonstrated in the images in the right half of Fig. 2. Note that the artefact removal process tends to mask slightly more than necessary, and this is the reason for some of the small masked regions visible in the figure. The number of views contributing data to each point is $n(\mathbf{r}) = \sum_i m_i(\mathbf{r})$. Together with $x_i(\mathbf{r})$, the quantities σ_i and $n(\mathbf{r})$ are also available to any subsequent data fusion process. In addition, a forward model [13] is used to predict the response, $E_{i|j}(\mathbf{r})$, to candidate defect, j, at any location in any view in the absence of microstructural noise. This provides further information for data fusion, potentially allowing prior knowledge about the type(s) of defect expected to be accommodated. Note that to obtain $E_{i|j}(\mathbf{r})$, the forward model output is scaled spatially in each view by $c_i(\mathbf{r})$ and by an overall factor in such a way that $E_{i|j}(\mathbf{r})$ is on scale consistent with $x_i(\mathbf{r})$. It is emphasised that all the data fusion methods described in the remainder of the paper are applied to multi-view image data that is generated through the same pre-processing steps. The differences between the data fusion approaches are only in how they combine the multi-view images.

3 DEVELOPMENT OF DATA FUSION METHOD

3.1 Definitions

This article is concerned with the binary detection problem: based on the information in an ultrasonic dataset, is a defect present or not in the component? There are therefore four possible outcomes:

- true positive – defect physically present and indication detected;
- false negative – defect physically present but no indication detected;
- true negative – defect physically absent and no indication detected;
- false positive – defect physically absent but indication detected.

For characterising the relative performance of different data fusion schemes two quantities are defined:

- Probability Of Detection (POD) – the fraction of physical defects that yield detectable indications;
- Probability of False Alarm (PFA) – the fraction of pristine datasets that will be wrongly flagged as containing defect indications.

Note that only the presence of a single defect in a dataset is considered here as (a) defects are generally rare and (b) this is regarded as the worst-case scenario for detection by the following logic. A first-order single-scattering model of multiple, ultrasonically-resolvable, weakly-scattering defects implies that ultrasonic responses do not interact and that the defect indications in multi-view images will superpose. Hence a weakly-scattering defect that is detectable in isolation is also detectable in the presence of other weakly-scattering defects. If one or more strongly-scattering defects is present than effects such as the shadowing of a smaller defect by a larger one become possible due to ultrasonic interactions between defects. This may mean that a defect detectable in isolation is no longer itself detectable in the presence of a larger defect. However, from the perspective of the overall detection problem, this does not matter as long as the larger defect is detected instead.

The output of applying a data fusion algorithm, γ, to multi-view images is a single image of a test statistic $z_\gamma(\mathbf{r})$. If the test statistic at any point in the ROI exceeds a given threshold, an indication is deemed to be detected. The relationship between POD and PFA as a function of threshold is plotted on a Receiver Operating Characteristic (ROC) curve. For a perfect detector, a threshold value can be found that enables a POD of unity and a PFA of zero to be achieved. However, in general, a compromise must be made between POD and PFA, and this depends on the requirements of an application. In NDE, the requirement is typically to achieve a specified POD; the lower the PFA at this POD, the better the inspection.

ROC curves are not altered if the test statistic considered is subject to a strictly monotonic mapping. To allow qualitative visual comparison of fused images, all test statistics used here are mapped in such a way that they are approximately homogeneous functions of input image amplitude.

3.2 Candidate data fusion methods

The data fusion methods investigated here can be divided into three types: novelty-detection methods where the only assumption is that a defect will give rise to a response that cannot be explained by a defect-free noise model; methods which are tuned to the detection of a specific defect based on prior knowledge of its response; and methods which can be tuned to the detection of multiple different defects. The fusion methods are described in the following subsections. Note that for brevity in this section, the dependence on position, $\mathbf{r}$, is omitted; in practice, a fusion method, γ, is applied independently at every point in the ROI to form the final fused image, $z_\gamma(\mathbf{r})$.

3.2.1 Novelty detection methods

Novelty detection methods do not use any prior knowledge of specific defect responses and only assume that potential defects result in responses that are different to the pristine case. Therefore, novelty detectors are binary classifiers where each point in the ROI is classed as being either pristine or not. The binary classification task can be regarded as a hypothesis test, where a suitable null hypothesis, H_0, for NDE applications is that the component is pristine. One methodology for hypothesis tests is based on p-values. A p-value is the probability that a value equal to, or more extreme than, the measured value would be obtained if H_0 is true. A low p-value therefore indicates that the null hypothesis is an unlikely explanation for the measurement and should be rejected, with the implication that a defect is potentially present. In the current application, the image intensities in each view (at a given spatial position) can be considered as different tests that can be converted into p-values and then combined. There is a considerable body of literature on combining p-values from multiple tests, most often in the context of obtaining an overall p-value when performing a meta-analysis of independent trials. Combining p-values has also been demonstrated previously as a successful data fusion approach for ultrasonic measurements in an NDE application [14].

Regardless of the method of p-value combination, the first step is to convert image intensities in each view into p-values. Under H_0 the artefact-free region of the ROI in the i^{th} view is a speckle pattern and the amplitude has a Rayleigh PDF with parameter σ_i,

$$\text{Rayleigh}(x|\sigma_i) = \frac{x}{\sigma_i^2}\exp\left(-\frac{x^2}{2\sigma_i^2}\right). \tag{3}$$

To convert the measured value, x_i, to a p-value, p_i, the PDF of H_0 is integrated from $x = x_i$ to ∞ in order to obtain the probability of obtaining a value equal to or more extreme than x_i under H_0. For a Rayleigh distribution this leads to a simple closed form expression

$$p_i = \int_{x_i}^{\infty} \text{Rayleigh}(x|\sigma_i)\, dx = \exp\left(-\frac{x_i^2}{2\sigma_i^2}\right). \tag{4}$$

There are many approaches to combine p-values, and the choice of which to use depends on the application [15][16]. In NDE, the primary requirement is to avoid type 2 errors (false-negatives), which is equivalent to missing defects, and the secondary requirement is to minimise type 1 errors (false-positives). On this basis, the method of combination should emphasise small p-values. Based on the recommendations in [15], Tippett's method [17] (referred to in [15] as the "minimum" method) and Fisher's method [18] (referred to in [15] as the "chi-square (2)" method) are considered here.

In Tippett's method the null hypothesis is rejected if any of the p-values is below a threshold, which is equivalent to defining a test statistic based on the minimum p-value:

$$p_T = \min_i p_i. \tag{5}$$

In order to obtain a test statistic for Tippett's method, z_T, that is a homogeneous function of input image amplitude, the mapping to a p-value can be reversed, resulting in:

$$z_T = \max_i \frac{x_i}{\sigma_i}. \tag{6}$$

Physically, fusion using Tippett's method means selecting (at each image position) the view where the signal amplitude relative to the noise is highest.

An alternative to Tippett's method is Fisher's method (also known Fisher's combined probability test), which considers the individual p-values as independent and multiplies them (through a summation of their logarithms) to obtain an overall p-value:

$$p_F = \mathrm{X}_{2k}^{-1}\left(-2\sum_{i=1}^{k} \ln p_i\right), \tag{7}$$

where $\mathrm{X}_{2k}^{-1}(\cdot)$ denotes the inverse of the Cumulative Distribution Function (CDF) of a χ^2 distribution with $2k$ degrees of freedom. Fisher's combined test can be considered as an analogue "AND" operation in terms of requiring all views to support the null hypothesis for it to be accepted. Equivalently, it can be regarded as an analogue "OR" operation from a defect detection point of view: rejection of the null hypothesis in any individual view leads to its overall rejection. As in the case of Tippett's method, it is preferable here to map the resulting p-value to a test statistic, z_F, that is a homogeneous function of input image amplitude:

$$z_F = \sqrt{\sum_{i=1}^{k} \left(\frac{x_i}{\sigma_i}\right)^2}. \tag{8}$$

Both Tippett's and Fisher's methods produce test statistics that measure departure from pristine case, H_0. For this reason, the fused images obtained from these methods have spatially uniform noise in the pristine case. However, their sensitivity to defects is not spatially uniform and also depends on the nature of the defect. This is an important consideration when comparing fused images between methods.

Neither Tippett's nor Fisher's methods require any information about the type of defect expected. Their relative performance as detectors depends on how defect indications are distributed among the views. Tippett's method is better when a defect is easily visible but only in a few views that are not known *a priori* [15] and therefore can be regarded as addressing the general case for data fusion, but not the edge case. On the other hand, Fisher's method also potentially addresses the edge case of weakly detectable defects that manifest in multiple views. Crucially, the performance of both is degraded by the inclusion of views that never contain defect indications as these views increase the possibility of false positives.

Thus, although neither Tippett's nor Fisher's method require explicit prior knowledge of defect responses, the choice of which views to include affects the performance of both. The selection of which views to include implicitly requires some prior knowledge about likely defect responses. For this reason, it is desirable to find a systematic means of including such knowledge.

3.2.2 Methods to detect single candidate defect

For methods that are tuned to the detection of a single candidate defect j (i.e. a defect of a specific type, size and orientation), the sensitivity, $E_{i|j}$, to that defect in each view must be known. If prior knowledge of a candidate defect is included in the fusion process, it is appropriate to spatially scale the fused image by the expected response to that defect; this justifies the application of a spatially-constant threshold to the fused image for detection of that type of defect. This contrasts with the fused images for the novelty-detectors described in the previous section, which have spatially uniform noise in pristine samples but non-spatially uniform sensitivity to defects.

In complex-valued, multi-view images (1) the defect response superposes algebraically with the microstructural speckle noise [19], [20]. However, when the intensity of an image is considered, the response from defect j in the presence of speckle noise has a Rician PDF [21],

$$\mathrm{Rice}\left(x|\sigma_i, E_{i|j}\right) = \frac{x}{\sigma_i^2}\exp\left[-\frac{\left(x^2 + E_{i|j}^2\right)}{2\sigma_i^2}\right] I_0\left(\frac{xE_{i|j}}{\sigma_i^2}\right), \tag{9}$$

where $I_0(\cdot)$ is a modified zero-order Bessel function of the first kind. The expected response, $\mu_{i|j}$, in view i to defect j is the mean of the Rician PDF

$$\mu_{i|j} = \sigma_i\sqrt{\frac{\pi}{2}}L_{1/2}\left(-\frac{E_{i|j}^2}{2\sigma_i^2}\right), \tag{10}$$

where $L_{1/2}(\cdot)$ is a Laguerre polynomial of order ½. For computational efficiency, the approximate expression given in Section S.2 in the supplementary material is used to compute the mean; this is accurate to within 4% of the exact expression above. If $E_{i|j}/\sigma_i$ is large, the expected response from the defect tends to $E_{i|j}$ (as the Rician distribution tends to a normal distribution), but for low $E_{i|j}/\sigma_i$ it tends to the mean of the noise Rayleigh distribution, $\sigma_i\sqrt{\pi/2}$, rather than zero. This point is important when the modified matched filter fusion method is developed below.

The simplest method for detecting a known defect (and the method that implicitly underlies many standard inspection techniques) is to use only data from the view with the highest signal to noise ratio based on the expected response: the 'best view'. The index, k_j, of the best view is therefore

$$k_j = \arg\max_i \frac{\mu_{i|j}}{\sigma_i}, \tag{11}$$

and the best-view test statistic is defined as:

$$z_{Vj} = \frac{x_{k_j}}{\mu_{k|j}}, \tag{12}$$

where the normalisation by $\mu_{k|j}$ ensures that the expected response from the best view is unity if defect j is present.

Note that the best view may not be constant throughout the ROI, and consequently the fused image will be composed of portions of different views. Like Tippett's method, this approach only exploits information from a single view; the difference is that the view to use here is defined *a priori* by the expected response of the candidate defect. To improve on this performance, information from multiple views needs to be combined in some manner.

One approach is to base the solution on the concept of a matched filter [22]. A matched filter is a weighted sum of inputs, with weights, w_i, chosen that maximise the signal to noise ratio of the result. In the classic matched filter development, each input channel is assumed to contain a known signal contribution that is polluted with additive Gaussian noise. If the inter-channel noise is assumed uncorrelated, it can be shown that the matched

filter weighting coefficient for a channel, w_i, is proportional to the expected amplitude of the signal of interest divided by the noise variance on that channel, i.e. $w_i = E_{i|j}/\sigma_i^2$, and the matched filter is then $\sum_i x_i E_{i|j}/\sigma_i^2$. It is instructive to compare the classic matched filter to Fisher's method (4). Both result in a summation of contributions from intensities; in Fisher's method the intensities are squared but in the matched filter they are not; both scale the contributions by the reciprocal of the noise variance in each view; and the matched filter further scales by the expected response to the candidate defect.

The classic matched filter works reasonably well in the current application, despite being incorrectly based on an additive Gaussian noise model. It can be shown [23] that in such cases the classic matched filter maximises SNR but does not necessarily result in an optimal detector. One way of formulating a matched filter expression is to start from a likelihood ratio test based on the PDFs of the signal and noise conditions. The PDFs for additive Gaussian noise (signal) and zero-mean Gaussian noise (noise) yield the classic matched filter. A better-matched filter for the current application can therefore be obtained if a Rician PDF (signal) and Rayleigh PDF (noise) are used in the formulation. The details of the derivation are provided in Section S.3 of the supplementary material. The result, termed the modified matched filter, is $\sum_i \ln I_0\left(x_i E_{i|j}/\sigma_i^2\right)$, where $I_0(\cdot)$ is a modified Bessel function of the first kind. A good approximation to $\ln I_0(v)$ is $\sqrt{v_0^2 + v^2} - v_0$ where the constant $v_0 = 2$. This approximation is within 4 % of the exact result and is how the modified matched filter is implemented here. If $x_i E_{ij}/\sigma_i^2 \gg v_0$, it can be seen that the approximation tends to $x_i E_{i|j}/\sigma_i^2$, and the modified match filter becomes the same as the classic matched filter. However, the crucial difference is that when $x_i E_{i|j}/\sigma_i^2$ is small the contribution to the summation from x_i is smaller than in the classical matched filter.

The test statistic based on the modified matched-filter principle is therefore

$$z_{Mj} = \frac{1}{\mu_j} \sum_i \left[\sqrt{v_0^2 + \left(\frac{x_i E_{i|j}}{\sigma_i^2} \right)^2} - v_0 \right], \tag{13}$$

where $v_0 = 2$, $E_{i|j}$ is the expected response in view i the presence of defect j, and

$$\mu_j = \sum_i \left[\sqrt{p_0^2 + \left(\frac{\mu_{ij} E_{i|j}}{\sigma_i^2} \right)^2} - p_0 \right]. \tag{14}$$

3.2.3 Methods to detect multiple candidate defects

In practice it is unlikely that an inspection is only concerned with the detection of one specific defect. Therefore, it is desirable to be able to address the general case of detecting any of $j = 1 \ldots J$ different candidate defects.

A naïve approach is to combine the sensitivities for the candidate defects into a mean sensitivity:

$$\bar{E}_i = \frac{1}{J} \sum_j E_{i|j}, \tag{15}$$

and use this in place of $E_{i|j}$ in the previous expressions for either the best view or modified matched filter method. The flaw of such logic is demonstrated in Section 3.3 below, using fusion based on the best view for mean sensitivity (termed mean best view) as an example. The flaw is that there is no guarantee of equal detection performance for all candidate defects. In an NDE scenario, it is generally the case that the POD for each candidate defect needs to exceed a certain level. This requirement suggests a more logical starting point for extending either the best view or modified matched filter approaches to test for multiple defects. The test statistics for each of those methods were defined previously such that the expected response from a given candidate defect, if present, was unity. Therefore, if multiple defects are of interest then a pragmatic solution is to test for each in turn and see if any of the responses exceeds a common threshold on a scale where unity represents the expected response. This gives the test statistics for the multiple best view method:

$$z_{MV} = \max_j z_{Vj}, \tag{16}$$

and for the multiple modified matched filter method:

$$z_{MM} = \max_j z_{Mj}. \tag{17}$$

In theory, a threshold set at unity for either test statistic would guarantee a POD of at least 50% for all candidate defects. (The POD of a candidate defect is theoretically exactly 50% in the test for that defect, but it may be higher if any of the responses below unity in that test are above unity in tests for other defects.)

Although the theoretical expected response of a given defect is unity in a test for that defect, the theoretical variance is not the same for all defects; qualitatively, the test statistics for defects that are weaker scatterers will exhibit a higher variance due to speckle noise. However, there is a more significant source of variance in practice due to the true morphology of a defect not matching $E_{i|j}$, since $E_{i|j}$ is predicted by a model of an idealised defect (and the model may involve other approximations). This variability is multiplicative rather than additive and has been found to be order ± 3 dB per view for the sensitivity model used here [13]. The pragmatic approach proposed is to assume that the shape of the PDF of the test statistics for multiple defects is the same for all candidate defects. In other words, a threshold applied to either the multiple best view or multiple modified matched filter test statistic is assumed to correspond to the same POD for any candidate defect.

Summary of candidate data fusion methods

Table 1 summarises the data fusion methods considered. In Section 3.3, the performance of the methods is compared using numerically simulated ROC curves.

Type	Name	Test statistic	Equation
No prior knowledge of defects	Tippett's method	z_T	(6)
	Fisher's method	z_F	(8)
Tuned to single defect	Best view	z_V	(12)
	Modified matched filter	z_M	(13)
Tuned to multiple defects	Mean best view	$z_{\bar{V}}$	(12, 15)
	Multiple best view	Z_{MV}	(16)
	Multiple modified matched filter	z_{MM}	(17)

Table 1 Summary of the fusion methods considered, associated test statistic symbols and equations.

3.3 Numerical comparison of candidate data fusion methods

3.3.1 Simulation of data for comparing fusion methods

A quantitative numerical comparison of the performance of candidate data fusion methods was performed prior to testing them on ultrasonic data. Example defect sensitivities, $E_{i|j}$, were produced using a single-frequency, ray-based forward model [13] for $J = 49$ candidate defects located at the centre of the ROI shown in Fig. 1(a). The 49 defects considered were: 7 circular cavities of diameter $0.5\lambda_L$, $0.75\lambda_L$, $1\lambda_L$, $1.25\lambda_L$, $1.5\lambda_L$, $1.75\lambda_L$ and $2\lambda_L$; 42 straight cracks for all combinations of length $0.5\lambda_L$, $0.75\lambda_L$, $1\lambda_L$, $1.25\lambda_L$, $1.5\lambda_L$, $1.75\lambda_L$ and $2\lambda_L$ and orientation 0°, 30°, 60°, 90°, 120° and 150°. The choice of these defects is to obtain a set of relative responses that might be encountered in a real inspection for illustrative purposes. Likewise, the relative noise levels, σ_i, in the different views were based on those measured experimentally on a copper sample in a previous paper [12]. These noise levels were increased by a factor of 40 in order to emphasise the difference between the fusion methods.

To obtain numerical estimates of data fusion performance, 200 realisations of defect responses, $X_{i|j}$, were generated for each of the $j = 1 \ldots J$ types of defect by drawing values from the appropriate Rician distribution (9), and 9,800 realisations of pristine responses, $X_{i|0}$, were generated by drawing values from a Rayleigh distribution (3). Each realisation is a vector $\{x_{i|j}\}^T$ of 21 image intensity values for either a defect ($j > 0$) or the pristine ($j = 0$) case. For each fusion technique, γ, a test statistic, $z_{\gamma|j}$, is computed for each realisation of each

case. For a given threshold level, α, the POD, $T_\gamma(\alpha)$, for each fusion technique is numerically estimated as the fraction of $z_{\gamma|j>0}$ that exceed α. The POD may be computed for a specific defect or by considering a population containing multiple types of defect. The PFA, $F_\gamma(\alpha)$, for each technique is numerically estimated as the fraction of responses, $z_{\gamma|j=0}$, from pristine realisations that exceed α. From these, the ROC curve, T_γ vs. F_γ, is computed. Note that in this section, the numerical data is drawn from PDFs that exactly match those assumed by the fusion processes.

3.3.2 Quantifying ROC curves

An ROC curve describes the performance of a detector. Hence one way to qualitatively compare the performance of candidate data fusion methods is to visually compare their ROC curves. A quantitative comparison requires a metric to be extracted from an ROC curve. A physically intuitive metric for NDE is the PFA at a specified POD level (e.g. PFA at 99% POD). However, because this depends on single point on an ROC curve it is rather sensitive when numerical or experimental data is considered. A more robust measure is to integrate the area under an ROC curve to obtain the Area Under Curve (AUC) metric:

$$A_\gamma = \int_0^1 T_\gamma \, dF_\gamma = \int_{-\infty}^{\infty} T_\gamma(\alpha) F'_\gamma(\alpha) \, d\alpha. \tag{18}$$

The AUC varies between 0.5 (detector cannot distinguish defects at all) and 1 (perfect detector). If the AUCs of multiple ROC curves are ranked, the ranking order will be the same as that for any other metric, provided the ROC curves do not cross.

3.3.3 Comparison results and discussion

Fig. 3 presents the results of the numerical simulation described above. The overall ROC curves shown in Fig. 3(a) summarise the performance of each data fusion method when applied to a population of data that contains defect responses from equal numbers of each of the 49 defect types. Note that this is an illustrative example based on the response of a defect at a single spatial location in the overall ROI; the actual ROC curve for a given defect in a real inspection would need to consider the possibility of that defect occurring at any spatial location in the ROI with the appropriate defect response for each location. Fig. 3(b) shows the AUCs for each individual type of defect (i.e. each AUC comes from a separate ROC curve computed for a population that contains only the response from a single type of defect).

First it is instructive to note from Fig. 3(a) the substantial improvement in performance going from the mean best view (in this case the T-T view, indicated by the black dotted line) to the other fusion methods. This demonstrates the general case for data fusion: essential inspection information is spread across multiple views, and no one view in isolation achieves anywhere near the same performance as any of the fusion methods. In order of increasing overall performance the other fusion methods can be seen to be multiple best view, Tippett's method, Fisher's method, and multiple modified matched filter. If the improvement going from the mean best view to multiple best view represents the general case for data fusion, the improvement going from the multiple best view to the multiple modified matched filter represents the edge case where a defect is barely visible in several views.

Consider the performance for populations of individual defects shown in Fig. 3(b). Cracks at 60° and 90° (horizontal) are the hardest to detect as there are no specular reflection paths in any view. These are the best defects in which to observe the spread of results. Fisher's method (green point-up triangles) outperforms Tippett's method (cyan point-down triangles). Likewise, the modified matched filter for a defect (red cross) outperforms the best view for a defect (black cross). The difference between both these pairs is down to the quantitative combination of information between views in Fisher's method and the modified matched filter. Again this is the edge-case where defects are at the limit of detectability in individual views, so quantitative combination of information improves performance.

The multiple best view and multiple modified matched filter techniques (black and red circles) generally perform worse than their single defect equivalents when applied to a population containing only one type of defect; this is because tests for defects not in the population tend to yield extra false positives.

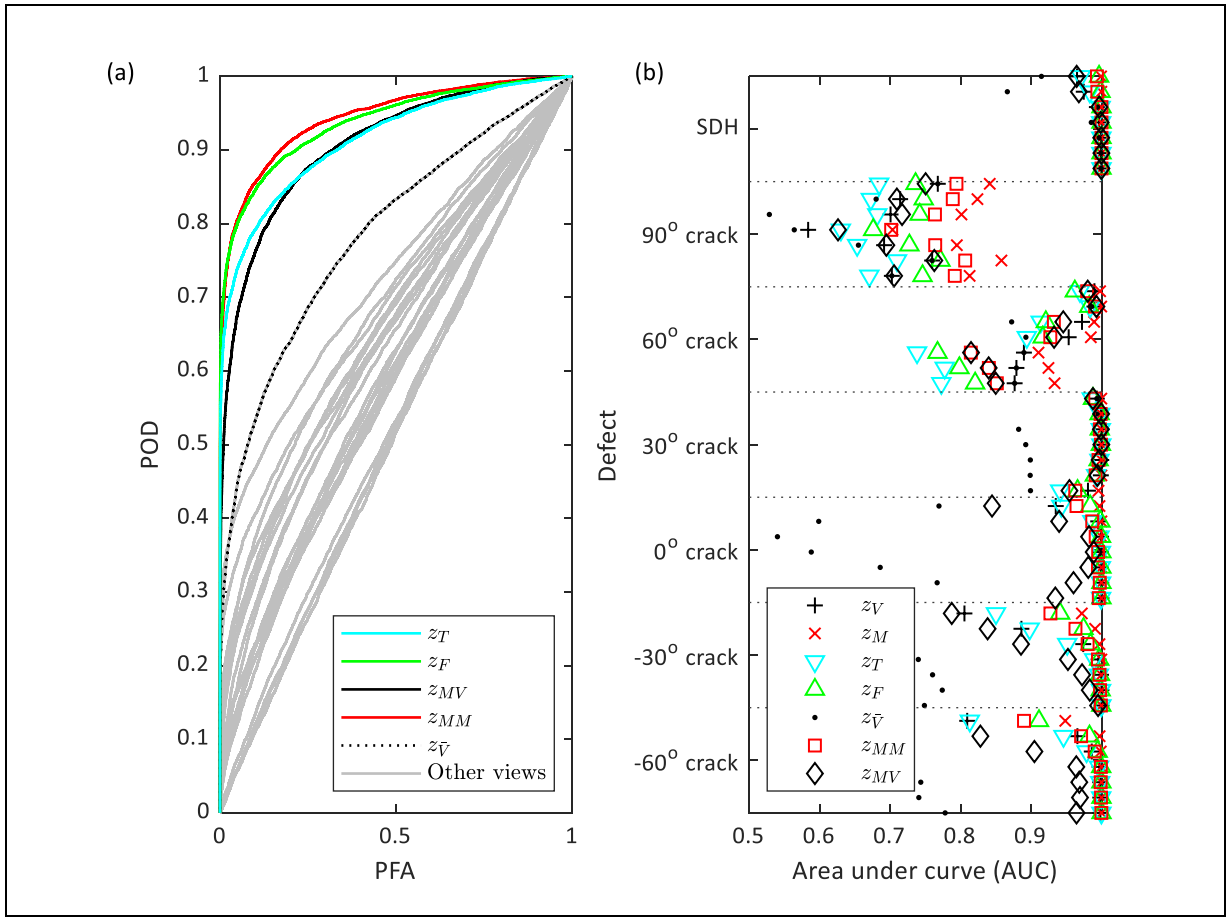

Fig. 3 Comparison of candidate data fusion methods using numerical data: (a) overall ROC curves for each fusion technique when applied to a population containing 49 different possible defects; (b) area under ROC curves for separate populations containing each individual defect. In (b) the defects of each orientation are ordered from largest (top) to smallest (bottom).

4 APPLICATION TO SYNTHESISED EXPERIMENTAL DATA

In the previous section, the relative performance of various fusion methods was compared by considering numerical data drawn from known distributions that exactly match those expected by the data fusion methods. The purpose of this section is to demonstrate and quantify the performance of the most promising fusion methods on realistic ultrasonic image data. The key difference in this section is that the measurement data is generated through numerical simulation of the complete ultrasonic process and thus comes from a distribution that is generally not exactly equal to the expected distribution. To perform this study requires a large population of independent datasets that are known to be pristine (i.e. defect-free) and a large population of datasets containing known defects. Obtaining the latter experimentally is extremely expensive and obtaining ground-truth data about real defects is also challenging. A pure simulation approach could be employed, but accurately capturing all sources of noise present in real measurements is difficult. For this reason, an approach that superposes simulated and experimental data is used that has been described and validated in other work [19], [20]. Details of the experimental, simulation and superposition methods are provided in Sections 4.1, 4.2 and 4.3 below.

4.1 Experimental measurements of noise

Using the configuration shown in Fig. 1(a), an array is scanned in the y direction over a region of a sample that is known to be defect free. As the array is scanned, $K_0 = 30$ datasets are recorded at y locations separated by 15 mm, a separation which has been determined [12] to be sufficient for the data from adjacent positions to be independent. An additional dataset is obtained in the presence of a Side-Drilled Hole (SDH) located at position $\mathbf{r_c}$ which is approximately in the centre of the ROI shown in Fig. 1(a). This provides a reference response that is

used in conjunction with the simulated response of the same target to obtain factors by which to scale the simulated results for each view prior to superposition with experimental data.

The procedure described in [12] is applied to the first pristine dataset to obtain the mask function, $m_i(\mathbf{r})$, levelling function, $c_i(\mathbf{r})$, and Rayleigh parameter for the noise, σ_i, for each view. Images, $I_i^{(Ek_0)}(\mathbf{r})$, for each view are generated for all $k_0 = 1 \ldots K_0$ pristine datasets. Likewise, images, $I_{i|ref}^{(E)}(\mathbf{r})$, for each view are generated from the reference dataset from a SDH.

In order to increase the number of datasets further, $K = 150$ images from pairs of pristine datasets are randomly superposed according to

$$I_i^{(Ek)}(\boldsymbol{r}) = \frac{1}{\sqrt{2}}\left(I_i^{\left(Ek_0^{(1)}\right)}(\boldsymbol{r}) + I_i^{\left(Ek_0^{(2)}\right)}(\boldsymbol{r})\right) \tag{19}$$

to yield $k = 1 \ldots K$ datasets for analysis, where $k_0^{(1)} \neq k_0^{(2)}$ are both drawn randomly from k_0. The factor of $1/\sqrt{2}$ means that the RMS noise in the result should be the same as that in the actual experimental data, assuming the noise at each point $\mathbf{r}$ is independent in datasets $k_0^{(1)}$ and $k_0^{(2)}$.

4.2 Finite element simulations of defect responses

Simulations are used to provide noise-free responses from known defects, which are then superposed onto the pristine experimental measurements described in Section 4.1. For the simulations, any suitable method can be employed and here time-domain Finite Element (FE) simulations are used; these capture all the physics of the scattering process and have the added advantage of being completely independent of the simulations used to obtain the sensitivity data for the fusion algorithms.

The FE simulations are performed using POGO [24], an explicit time-marching code that is designed to exploit the computational power of Graphics Processing Units (GPUs). The modelling domain is 2D and square linear plane-strain elements with side length 0.02 mm were used, with a time step of 1 ns. A limitation of the current version of POGO is that it does not support liquid-solid coupling. For this reason, the POGO modelling domain is a 2D model of the test piece in vacuum and the following procedure is used to obtain equivalent immersion-coupled data from the model. First the time-dependent ultrasonic pressure from an element, *T*, in an array radiating into liquid is simulated analytically at points in the liquid, $\mathbf{r}_s$, at the location of the test piece surface using the following far-field, narrow-band approximation:

$$P_T(\mathbf{r}_s, t) = \sum_{i=1}^{n} \frac{s\left(t - \frac{\left|\mathbf{r}_T^{(i)} - \mathbf{r}_s\right|}{c_w}\right)}{\sqrt{\left|\mathbf{r}_T^{(i)} - \mathbf{r}_s\right|}} \tag{20}$$

where $s(t)$ is the time-domain input signal to the array, c_w is the speed of sound in the liquid couplant, and $\mathbf{r}_T^{(i)}$ are the location of $n = 3$ equi-spaced points across the width of the transmitting element, the combined fields from which are used to capture the element directivity. The pressure is converted into time-dependent out-of-plane forces that are applied at the nodes on the surface of the test piece in the FE model. The time-dependent out-of-plane displacements, $u_z^T(\mathbf{r}_s, t)$, at the surface nodes are recorded and used as sources in an analytical model of radiation into an infinite liquid. The superposition of the fields from all such sources at appropriate positions in the liquid provides the received signals at $n = 3$ points on each array element which are summed to provide the received time-domain signal from that element:

$$u_{TR}(t) = \sum_{i=1}^{n} \int \frac{u_z^T\left(\mathbf{r}_s, t - \left|\mathbf{r}_R^{(i)} - \mathbf{r}_s\right|/c_w\right)}{\sqrt{\left|\mathbf{r}_{\boldsymbol{R}}^{(\boldsymbol{i})} - \mathbf{r}_s\right|}} d\mathbf{r}_s \tag{21}$$

The procedure is repeated using each element in turn as a transmitter, executing the FE simulation, and recording the time history at each receiving array element. In this way the complete FMC dataset is simulated. The signals, $u_{TR}(t)$, are again converted to analytic signals, $h_{TR}(t)$, prior to imaging using a Hilbert transform to generate the imaginary quadrature component.

The output of the FE simulation for the j^{th} defect, $h_{TR|j}(t)$, is processed using (1) to obtain the images $I_{i|j}^{(S)}(\mathbf{r})$ for each view. A defect-free simulation is performed to obtain images $I_{i|0}^{(S)}(\mathbf{r})$ that contain only responses from geometric features. Finally, a simulation of the response of a SDH at location $\mathbf{r_c}$ is also performed to obtain reference multi-view images, $I_{i|ref}^{(S)}(\mathbf{r})$. In total, FE simulations were used to produce raw datasets for 11 different defects as listed in Table 2. The cracks at orientations of -20° were chosen to be representative of fusion face cracks on the near face of a 20° single-V butt weld and the 0° cracks to be representative of inter-run defects within a weld.

Number	Type	Size (mm)	Orientation (° from vertical)	(*x*, *z*) position in ROI (mm)
1	Circular void	1.0	-	(6.8, 11.0)
2	Crack	5.0	-20	(6.8, 11.0)
3	Crack	2.0	-20	(6.8, 11.0)
4	Crack	1.5	-20	(6.8, 11.0)
5	Crack	1.0	-20	(6.8, 11.0)
6	Crack	0.5	0	(6.8, 4.6)
7	Crack	1.0	0	(6.8, 4.6)
8	Crack	1.0	0	(6.8, 17,9)
9	Crack	1.0	0	(6.8, 11.0)
10	Crack	1.0	-20	(4.5, 4.6)
11	Crack	1.0	-20	(9.3, 17.9)

Table 2 Defects modelled in FE. The longitudinal and shear wavelengths at the 5 MHz centre frequency of the simulations were 0.94 and 0.45 mm respectively and the position coordinates are relative to the image origin defined in Fig. 1(a).

4.3 Superposition

Although the FE model captures the scattering process at the defect, it does not include material attenuation and models the boundaries of the component as free rather than immersed in water. For this reason, the simulated data in each view needs to be scaled separately before superposing the experimental data. Furthermore, the output of each FE simulation contains both the defect response and the effect of ultrasonic interactions with the geometry of the component. However, the latter are already captured in the experimental pristine data and so need to be removed from the simulated data prior to superposition. The overall superposition process can thus be written:

$$I_{i|j}^{(k)}(\mathbf{r}) = I_{i}^{(Ek)}(\mathbf{r}) + \beta_i \left[I_{i|j}^{(S)}(\mathbf{r}) - I_{i|0}^{(S)}(\mathbf{r}) \right] \tag{22}$$

The view-dependent scale factor, β_i, obtained from the reference SDH:

$$\beta_i = \beta \left| \frac{I_{i|ref}^{(E)}(\mathbf{r_c})}{I_{i|ref}^{(S)}(\mathbf{r_c})} \right| \tag{23}$$

where β is an overall scale factor that allows the SNR to be altered across all views if desired. The accuracy of the approach is demonstrated in Fig. S3 in Section S.4 of the supplementary material.

Finally, the corrected response in the i^{th} view for the j^{th} defect and the k^{th} realisation of experimental noise is obtained from:

$$x_{i|j}^{(k)}(\mathbf{r}) = c_i(\mathbf{r})m_i(\mathbf{r})\left|I_{i|j}^{(k)}(\boldsymbol{r})\right|. \tag{24}$$

This forms the input into data fusion algorithms.

4.4 Detection results

The results of applying the various data fusion algorithms to the synthesised data are plotted in Fig. 4. Here the amplitude of the noise has been increased by a factor of $\beta = 10$ to emphasise the differences between the data fusion methods. The ROI is as defined in Fig. 1(a). The mean best view, multiple best view and multiple modified matched filter methods require sensitivity maps for each candidate defect at every position in each view. These maps are generated completely independently of the FE simulations again using the single-frequency, ray-based sensitivity model [13]. The library of candidate defects used for the fusion methods that require one contains 1 mm long cracks at angles of 0°, -10°, -20° and -30° to the vertical, thus spanning the angular range of the defects present in the simulated data, but not the length range.

The graphs in Fig. 4 show ROC curves computed for a population containing 150 datasets for each of the 11 defects listed in Table 2, together with 150 pristine datasets containing experimental noise (i.e. a total population of 1,800 datasets). The ROC curves in Fig. 4(a) are broadly consistent with the results in the previous section: in general the multiple modified matched filter (red line) performs best, the multiple best view method (black line) is worst, with Fisher's and Tippett's methods in between. For cracks of increasing size at the same orientation and position (e.g. defects 5, 4, 3, and 2), it can be seen in Fig. 4(b) that the performance of all methods either improves with crack size or, in the case of the multiple matched filter, is close to unity for all sizes; this is despite the multiple matched filter and best view methods only testing for the smallest of the crack sizes present. In contrast to the previous results in Fig. 3, the mean best view method performs better than the multiple best view method when applied to a population of different defects. This is because in Fig. 4 the range of defects (specifically crack orientation angle) in the library and the test population are quite small. The mean best view based on the mean response of defects in the library therefore corresponds to a view that provides reasonable performance for most of the defects in the real population.

Note that the ROC curves shown in Fig. 4 are based on the POD of defects at the specific locations listed in Table 2, rather than the POD for a given type of defect throughout the whole ROI. By contrast, the PFA is based on indications that can occur at any location in the ROI. This inconsistency in the POD and PFA definitions means that the ROC curves (and associated AUC metrics) do not always provide a fair comparison between the different data fusion methods. This is particularly relevant for the multiple best view and multiple matched filter methods. For these methods, the output is scaled to provide uniform sensitivity throughout the ROI, with the intention of increasing the POD for defects of a given type occurring at any location. This is accompanied by an increase in PFA because of the amplification of weak signals in regions of the ROI where the expected defect response is small. However, while the ROC curves in Fig. 4 capture the increase in PFA, the expected increase in POD is not revealed because the POD is only for a defect at a single location. This is one reason why the multiple matched view appears to perform worse than the mean best view; the other reason is that the range of defects (specifically crack orientation angle) considered is quite small so the mean best view in this case leads to a fairly a good view for detecting most defects.

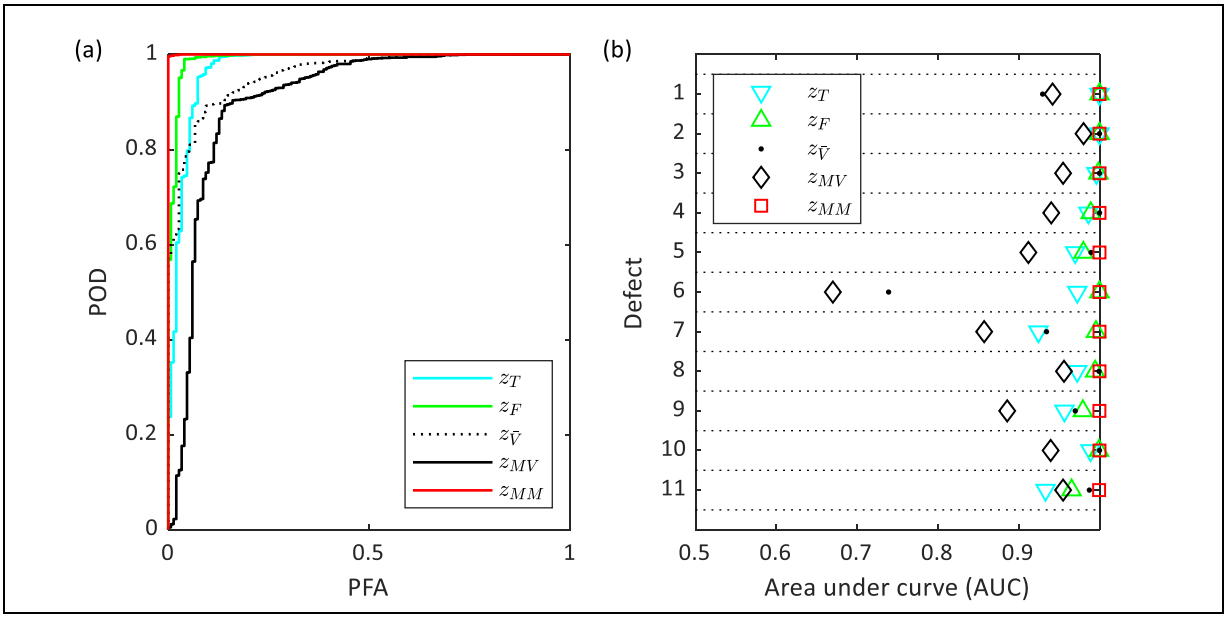

Fig. 4 (a) ROC curves obtained from merged experimental and finite element data for various fusion methods and (b) the AUC for populations of individual defects. The library of candidate defects for the methods requiring one contained 1 mm long cracks inclined relative to vertical at 0°, -10°, -20° and -30°.

As the output of the various data fusion algorithms are fused images, it is instructive to view some examples of these. Fig. 5 shows examples of the fused images for the 5 fusion methods using one randomly-selected dataset for each of the 11 defect types. As in Fig. 4, the defect libraries for the methods requiring prior information only contain 1 mm long cracks at 0°, -10°, -20° and -30° to the vertical. The spatial uniformity of noise in the novelty detection methods and its lack of uniformity in the other methods is evident. It should be stressed that visual comparison of images between fusion methods can be misleading. Although the test statistics are designed as far as possible to be homogeneous functions of the original ultrasonic image amplitudes, the noise distribution in the fused images depends on the fusion method.

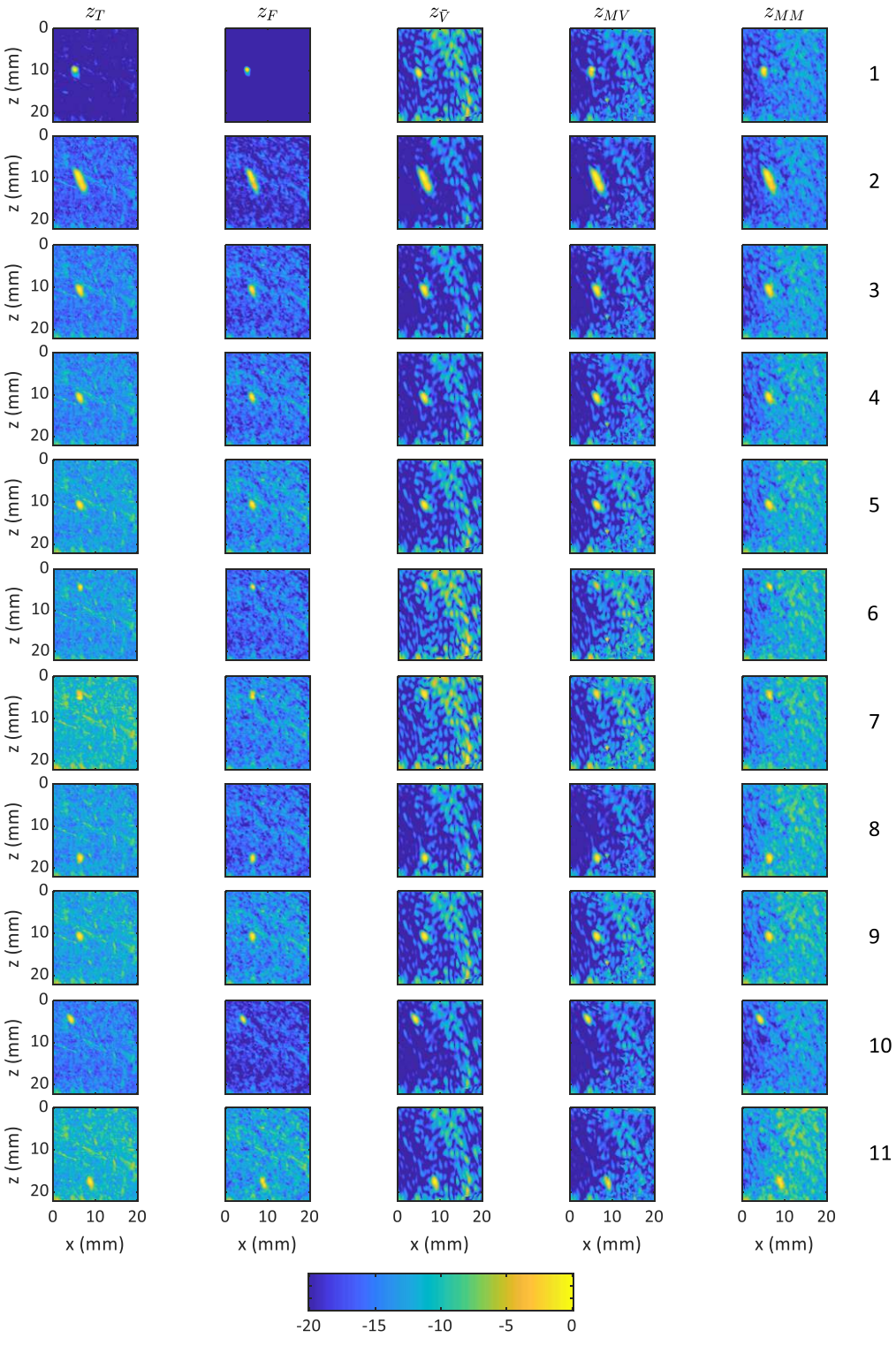

Fig. 5 Example fused images for 5 fusion methods applied to one randomly-selected dataset for each of the 11 defects modelled. For methods requiring libraries of candidate defects the libraries contained 1 mm long cracks inclined at 0°, -10° and -20° to the vertical only. The colour scale is -20 to 0 dB relative to the peak response in each fused image.

5 APPLICATION TO EXPERIMENTAL DATA

5.1 Experimental procedure

A weak signal from the FBH is visible in Fig. 2 in several views (L-L, LL-L and LL-T), but here the array has been optimally aligned so that the axis of the FBH is in the imaging plane. Consequently, specular reflections from the tip of the FBH occur in the imaging plane. A demonstration of the improvement in detection performance that data fusion provides in a realistic inspection scenario can be obtained by recording data as the array is translated past the defect in the y-direction in small increments. This mimics what happens in a real scan, in that individual defects may not lie exactly in the imaging plane at any one scan position or they may be oriented such that their specular reflection directions are not in the imaging plane.

To achieve this, a computer-controlled mechanical scanning device is used to move the array in 0.5 mm increments in the y-direction from -15 mm to 15 mm with respect to the axis of the FBH. In the $x - z$ plane, the position of the array relative to the sample is kept constant in the configuration shown in Fig. 1(a). At each position, an FMC dataset is recorded, multi-view images are generated, and the results fused using either Fisher's method, the best view method or the modified matched filter. For the latter two, the candidate defect is a 1 mm long vertical crack. Because the sensitivity model is based on a 2D defect (i.e. the candidate crack is 1 mm long in the z-direction and infinitely long in the y-direction) the sensitivity model response is much larger than that of the physical defect represented by the ∅1 mm tip of the FBH in the experimental sample. However, it is expected that the angular scattering response of the ∅1 mm tip of the FBH in the $x - z$ plane will be similar to that of a 1 mm long 2D crack, but that its response will be significantly weaker due both its smaller scattering cross section and the fact that the scattered waves diverge in the y dimension as well as the imaging plane. For this reason, the candidate defect amplitude was scaled down by a factor of ten in all views; this factor was estimated based on a comparison of sensitivity model prediction with the measured experimental response in the T-T view at the $y = 0$ position.

5.2 Detection results

The results of applying data fusion to this experimental data are presented in Fig. 6. It is not possible to present the detection results as conventional ROC curves as it is not known *a priori* from which y-positions the defect should be detectable (i.e. the denominators in the POD and PFA fractions are unknown). Instead, the detection results are presented as free response curves in Fig. 6(a). These are calculated as follows. A 5 mm square detection box is defined in the $x - z$ imaging plane centred on the known location of the defect. For each dataset, the peak values of the fused result inside and outside the detection box are obtained. For a given detection threshold, the number of apparent true positives is the number of datasets where the peak value inside the detection box exceeds the threshold, and the number of apparent false positives is the number of datasets where the peak value outside the box exceeds the threshold but the peak value within the box does not.

The detection threshold is swept to obtain the free response curves shown in Fig. 6(a) for each fusion method. All three methods can detect the defect in multiple datasets with zero apparent false positives; the threshold at which this occurs is defined as the zero-false-positive threshold. As the detection threshold is decreased the number of apparent true positives rises slightly, while the number of apparent false positives rises steadily. For each fusion method, there is a detection threshold below which no further changes can occur; below this threshold, detection just depends on whether the peak value for a method occurs within the detection box (apparent true positive) or outside (apparent false positive). The ROC curves terminate at this point, which for each curve lies on the line where the sum of apparent true positives and false positives equals the total number of datasets. Throughout Fig. 6(a) the modified matched filter method detects defects from more positions than the other methods.

The detection range in the y-direction for each fusion method is shown in Fig. 6(b) where the peak amplitudes within the detection box and outside the box are plotted as a function of y position. In this graph the amplitudes are shown in dB relative to the zero-false-positive threshold for each fusion method. The solid coloured bars above the graphs indicate the range of y over which the defect is detected with zero false-positives (i.e. the defect response exceeds the zero-false-positive threshold). Figs. 6(c) and (d) show example fused images obtained at $y = 1$ and $y = 6$ mm, indicated by the vertical blue dashed and dash-dot lines in Fig. 6(b). These images are presented over a 20 dB dynamic range, again with 0 dB set at the zero-false-positive threshold. In each case the upper end of the range is set to the maximum value obtained at all y-locations for each fusion

technique, with any values over 0 dB highlighted in red. The red regions in images therefore represent indications from the FBH that could be detected with zero-false-positives at any y position. The results at $y = 1$ mm in Fig. 6(c) are when the array is close to the centreline of the FBH and it can be seen that it is readily detectable by all fusion methods. The results at $y = 6$ mm in Fig. 6(d) are when the FBH is on the limit of detectability. Although an indication is visible at the correct location for all fusion methods, the amplitude of the defect response in the best view method fails to exceed the zero-false-positive threshold for that method. The increase in detectable range obtained using the modified matched filter method (and to a lesser extent Fisher's method) over the best view is an experimental demonstration of the edge case benefit of data fusion.

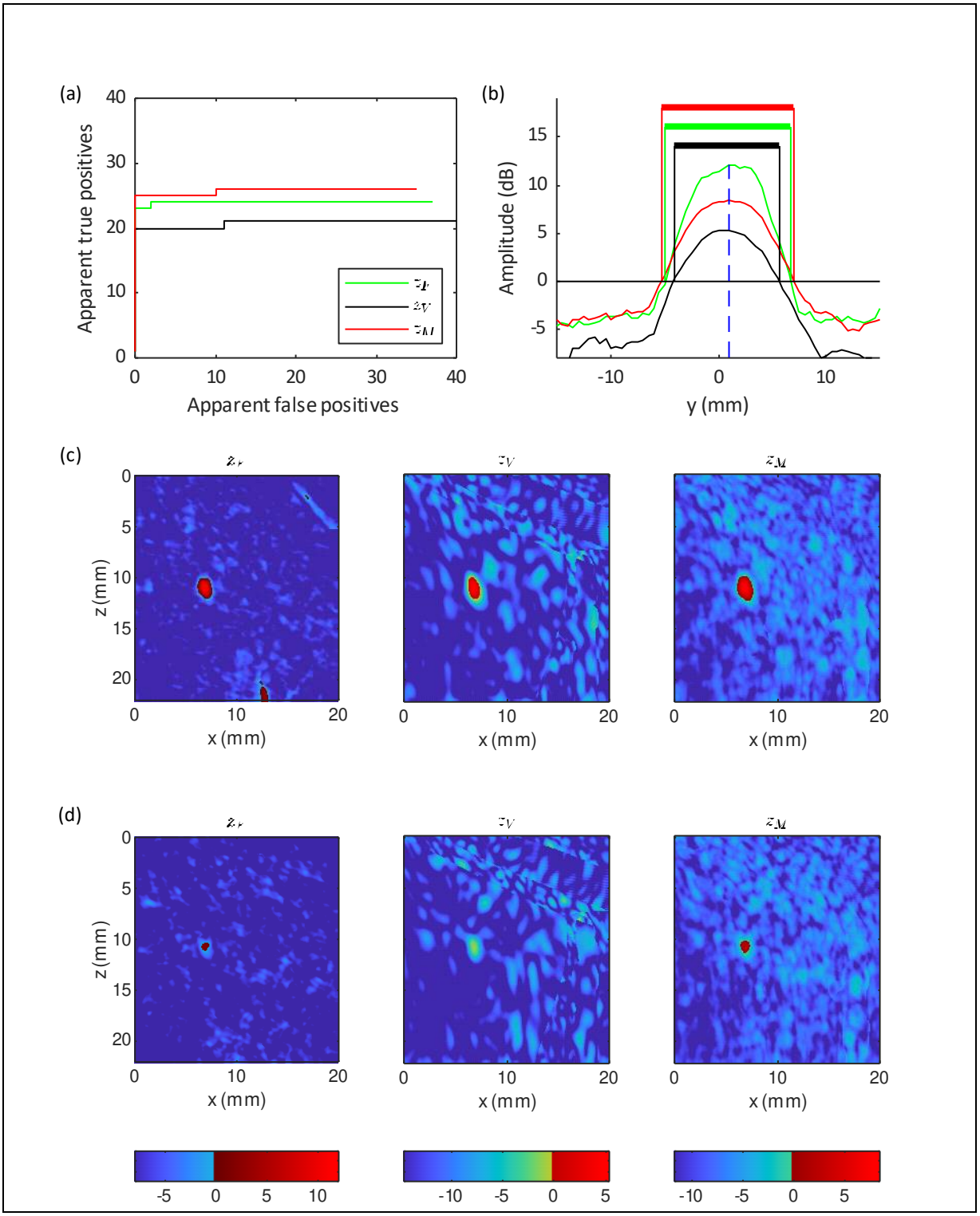

Fig. 6 (a) Free ROC curves for experimental data; (b) amplitude of response at defect location as a function of y-position; (c) example fused images from the dataset obtained at the position $y = 1$ mm indicated by the vertical blue dashed line in (b); (d) example fused images from the dataset obtained at the position $y = 6$ mm indicated by the vertical blue dash-dot line in (b). In (b-d), amplitudes are expressed in dB relative to the zero-false-positive level for that fusion technique. The solid horizontal bars in (b) indicate the ranges of y where the defect is detectable with zero false-positives. In (c) and (d) the fused images are plotted over a 20 dB dynamic range with the upper limit set to the maximum value recorded at any y position for each method. Values over 0 dB for each method are shown in red; to highlight indications from the FBH which can be identified without false positives at any y position.

6 DISCUSSION

6.1 Dealing with uncertainty and mis-registration of views

For all data fusion methods, it is necessary to experimentally calibrate the noise and locate any imaging artefacts. This can be performed [12] using a single FMC dataset acquired from a pristine region. The position of imaging artefacts is extremely sensitive to parameters such as array stand-off and angle. Hence if these cannot be controlled precisely during a scan, the masked regions describing potential artefact locations obtained from the pristine dataset should be dilated using suitable image processing algorithms. This will be at the expense of reducing the number of views able to contribute to data fusion in the vicinity of artefacts.

It has been assumed here that the multi-view images are near perfectly co-registered and great care has been taken in the experimental measurements presented to ensure that this is achieved. In real inspections, such precision may not always be possible, and some degree of mis-registration is to be expected due to imprecisely known material properties or geometry. Fusion using Tippett's or the (multiple) best view method does not rely on quantitative combination of values from multiple views to form the fused image at any one point. For these methods the detection performance is likely to be resilient to mis-registration between views. However, fusion using Fisher's or the (multiple) modified matched filter method relies on the quantitative combination of values from multiple views. Here mis-registration is more important and a possible mitigation is to use resolution elements or resels [14], which are small regions of a pre-defined size, that represents the co-registration accuracy. Here the values used for fusion, $x_i'(\boldsymbol{r})$, at a nominal spatial location, $\mathbf{r}$, in each view are not taken as the normalised image intensity at precisely that location, $x_i(\mathbf{r})$, but rather as the maximum normalised image intensity within a resel, i.e. $x_i'(\mathbf{r}) = \max x_i(\mathbf{q})$ for $|\mathbf{q} - \mathbf{r}| \leq \delta/2$ where δ is the size of a resel (here assumed circular, although square ones may be more practical computationally). The use of resels to deal with mis-registration reduces the resolution of the final fused images but also reduces detection performance. This is because in pristine regions the maximum normalised intensity in a resel has a different statistical distribution to the Rayleigh distribution of intensity at individual points. Calculation of the maximum intensity distribution in a resel is not straightforward as it is a function of the spatial correlation of image values as well as resel size. However, it is evident that on average the maximum intensity within a resel must be higher than the average intensity at a single point and will increase with resel size. For this reason, every effort should be made to minimise the misregistration between views if using resels.

6.2 Computational cost

Prior to performing an inspection, it is necessary to compute sensitivity maps for each defect in each view if the data fusion method is one tuned to particular defects. The single-frequency method proposed in [13] enables good approximations of such maps to be produced very rapidly (order of one minute per candidate defect) for a given inspection geometry, and the inputs to the sensitivity model are the scattering matrices of canonical defects, which only need to be computed once.

When an inspection is being performed, the main computational burden is producing the input multi-view images. The speed at which this can be performed is increasing all the time, as the imaging operation can be parallelised, e.g. on PC-based GPUs or in hardware-based FPGAs. Consequently, it is unlikely to be a rate-determining step for multi-view imaging in the future. The computation required for the fusion itself is relatively modest compared to the image formation stage: the Tippett and Fisher methods require the multi-view images be normalised to the local noise and then either the maximum normalised image taken at each point (Tippett) or the normalised image values squared and summed (Fisher). The modified matched filter requires a weighted sum across views for each candidate defect prior to taking the maximum across all candidate defects. For the fusion methods for multiple defects, the underlying fusion calculation needs to be repeated for each candidate. This may lead to a high computational load if the number of candidate defects is large. However, there is scope for reducing the number of candidate defects that actually need to be considered: if the test for one candidate defect can be shown to always leads to a larger test statistic than another, only the test statistic for the weaker one needs to be computed from an experimental measurement. For example, the modified matched filter is a sum of nonlinear but monotonic functions of $x_i E_{i|j}/\sigma_i^2$. Consider two candidate defects, $j = 1$ and $j = 2$. If $E_{i|2}$ is found to be greater than $E_{i|1}$ for every view, then the test statistic z_{M2} will always be greater than z_{M1} when the modified matched filters for the two candidate defects are applied to the same set of measurements $\{x_i\}$. There is therefore no need to test for candidate defect $j = 2$ since if z_{M1} exceeds the detection threshold, so will z_{M2}. In practice this is likely to be the case for defects with increasing size of the same type and orientation.

6.3 Operator assistance vs. inspection automation

Multi-view imaging generates a vast amount of data. The increased data provides more information than a single view but places a very high burden on operators. Consequently, one application of data fusion is to reduce operator burden by enabling operators to focus their attention just on datasets containing potential defect indications. In this case, there does not need to be a hard detection threshold placed on the data fusion output. Instead, a possible protocol is to place the peak data fusion outputs from each dataset in order from largest to smallest. The operator then works down the list, investigating each dataset in turn until it is deemed that pristine datasets have been reached. In this use case, any of the proposed data fusion methods could be used, as detection is not based on the data fusion output itself. In the absence of any prior knowledge about the type of defects present, a single modified matched filter based on the scattering from a small void is the recommended approach; in contrast to the anomaly-detection methods (Tippett's and Fisher's) this naturally addresses the problem of determining which views to include.

Complete inspection automation requires the data fusion process itself to determine if a defect is present. For this purpose, fusion methods that just detect anomalies are inadequate; there needs to be a direct link between the data fusion output and the presence of a defect so that a detection threshold can be set. This relies on sensitivity maps, which must be scaled according to a known reference reflector (e.g. the component back wall) in the experimental data. It would also be prudent for the inspection protocol to include a calibration check on a test-piece containing one or more known defects (e.g. side-drilled holes), which would provide confidence in the spatial accuracy of the sensitivity maps within the inspection volume. The choice of fusion method for automation depends on the inspection target. For the general case, the (multiple) best view method is straightforward and expected to be more tolerant to mis-registration between views. For the edge case of detecting very small defects, the (multiple) modified matched filter method is recommended, but the full benefit will only be achieved if the mis-registration between views is small.

7 CONCLUSION

It has been shown that the fusion of multi-view images obtained from ultrasonic arrays can improve defect detection performance. Specifically:

1. The performance improvements due to data fusion are for two different reasons corresponding to the general and edge cases. The general case is when no one view enables all defects of interest to be detected, even though individual defects can be readily detected if the appropriate view for that defect is used. The edge case is improved performance for defects on the limit of detectability, which are barely visible in any view.
2. Fusion can be performed without explicit knowledge about the defects to be detected. Fisher's and Tippett's methods for anomaly detection achieve this and generally produce better detection performance than that of a single view in isolation when applied to a population of different defects. However, the choice of the views to use in these methods is subjective and affects their performance: too few may result in useful views being ignored and too many may result in the inclusion of views that are not useful for detection and only generate false positives. Determining the views to include in Fisher's and Tippet's methods implicitly requires knowledge of potential defect responses.
3. In the absence of other information, a small (sub-wavelength) void could be used as a representative defect from which an expected response map for each view could be generated. This would provide a tool for ranking views in terms of the expected response:noise ratio and deciding which to include in Fisher's and Tippet's methods. However, this still requires a decision to be made about where to make the expected response:noise ratio cut-off. For this reason, once a sensitivity map is obtained, it is preferable to use one of the tuned data fusion methods that exploit this information directly, as summarised in the following points.
4. The best view and modified matched filter methods use expected response:noise ratio information for a specified defect. The best view method simply selects data from the view with the highest expected response:noise ratio at each spatial location but does not fuse data from multiple views at individual locations. The modified matched filter method combines contributions from all available views with consideration of the expected response:noise ratio in each view. This method is shown to outperform other methods when applied to individual defects of the type expected.
5. When multiple types and orientations of defects need to be detected the best view and modified matched filter approaches can be used to test for each candidate defect in turn. Both methods return

test statistics for different candidate defects on a consistent scale that relates to their PoD. For this reason, it is justifiable to apply a detection threshold at the same level across the tests as a way of achieving a consistent PoD for all candidate defects.

6. The multiple best view method is expected to be more resilient to mis-registration than the multiple modified matched filter as it does not quantitatively combine information across multiple views. It does not have as good performance in edge cases as the multiple matched filter method, but if the motivation for performing data fusion is the general case then the multiple best view method is a robust choice.
7. The multiple modified matched filter method is the most appropriate for edge case applications where the defects of interest are at the limits of detectability in any view, but it does require accurate co-registration of the different views in order to achieve its best performance.

The first use case for multi-view data fusion is likely to be reducing operator burden. Even a scan around a girth weld in a 12" diameter pipe may generate over 20,000 images, and in many cases, there will be no defects. To manually examine the images for defect indications would be extremely time consuming, even if the defects of interest are expected to give clear indications in one or more views. Data fusion provides a systematic method for sorting the datasets in order of severity of defect indications present. The operator can then work down the list from the most severe indication until they are satisfied that weaker indications are not due to defects, a process which in most cases might only require detailed examination of the first few datasets. This is an example of practical application of data fusion in the general case.

The longer-term use case for data fusion is to reliably extend the performance of an inspection to detect smaller defects. This is the edge case, which is extremely important because it defines the limit of ultrasonic NDE capability. An extension in NDE capability has wider benefits, for example by enabling longer intervals between inspections or reduced design conservatism in engineering components.

Finally, it is noted the data fusion methods described here all fuse probabilistic rather than physical quantities (i.e. p-values for Tippett's and Fisher's methods; likelihood ratios for best view and matched filter methods). This suggests that they could form the starting point for fusing heterogeneous physical quantities acquired across different NDE modalities (e.g. ultrasonic, radiographic, eddy current and thermographic techniques), if the physical data is first converted to homogeneous probabilistic data. However, consideration would also need to be given to the compatibility of the spatial regions (shape, dimensionality and resolution) represented by the probabilistic values from different modalities prior to fusion.

ACKNOWLEDGEMENTS

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) [grant numbers EP/N015924/1, EP/N015533/1]; the UK Research Centre in NDE (RCNDE); BAE Systems; EDF Energy; Hitachi; and Wood Group (formerly Amec Foster Wheeler UK).

DATA ACCESS STATEMENT

All data and code is available for download from the Research Data Repository of University of Bristol at [DOI will be added if paper accepted].

Temporary copy of data and code for review is available at:

https://www.dropbox.com/sh/eic8xd4v3s0okby/AADsN9rZ8_J9-Xqy1VCBAglUa?dl=0

REFERENCES

[1] B. W. Drinkwater and P. D. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," *NDT E Int.*, vol. 39, no. 7, pp. 525–541, 2006, doi: 10.1016/j.ndteint.2006.03.006.

[2] C. Holmes, B. W. Drinkwater, and P. D. Wilcox, "Post-processing of the full matrix of ultrasonic transmit–receive array data for non-destructive evaluation," *NDT E Int.*, vol. 38, no. 8, pp. 701–711, 2005, doi: 10.1016/j.ndteint.2005.04.002.

[3] N. Pörtzgen, D. Gisolf, and G. Blacquière, "Inverse wave field extrapolation: A different NDI approach to imaging defects," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 54, no. 1, pp. 118–127, 2007, doi: 10.1109/TUFFC.2007.217.

[4] A. J. Hunter, B. W. Drinkwater, and P. D. Wilcox, “The wavenumber algorithm for full-matrix imaging using an ultrasonic array,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 55, no. 11, pp. 2450–2462, 2008, doi: 10.1109/TUFFC.952.

[5] L. Merabet, S. Robert, and C. Prada, “Comparative study of 2D ultrasound imaging methods in the f-k domain and evaluation of their performances in a realistic NDT configuration,” in *AIP Conference Proceedings*, 2018, vol. 1949, doi: 10.1063/1.5031654.

[6] J. Zhang, B. W. Drinkwater, P. D. Wilcox, and A. J. Hunter, “Defect detection using ultrasonic arrays: The multi-mode total focusing method,” *NDT E Int.*, vol. 43, no. 2, pp. 123–133, 2010, doi: 10.1016/j.ndteint.2009.10.001.

[7] L. Le Jeune, S. Robert, E. Lopez Villaverde, and C. Prada, “Plane Wave Imaging for ultrasonic non-destructive testing: Generalization to multimodal imaging,” *Ultrasonics*, vol. 64, pp. 128–138, 2016, doi: 10.1016/j.ultras.2015.08.008.

[8] X. L. Han, W. T. Wu, P. Li, and J. Lin, “Combination of direct, half-skip and full-skip TFM to characterize multi-faceted crack,” in *2015 IEEE International Ultrasonics Symposium, IUS 2015*, 2015, doi: 10.1109/ULTSYM.2015.0339.

[9] L. Hörchens, X. Deleye, and K. Chougrani, “Ultrasonic imaging of welds using boundary reflections,” in *AIP Conference Proceedings*, 2013, vol. 1511, pp. 1051–1058, doi: 10.1063/1.4789159.

[10] K. Sy, P. Brédif, E. Iakovleva, O. Roy, and D. Lesselier, “Development of the specular echoes estimator to predict relevant modes for Total Focusing Method imaging,” *NDT E Int.*, vol. 99, pp. 134–140, 2018, doi: 10.1016/J.NDTEINT.2018.07.005.

[11] R. F. Wagner, S. W. Smith, J. M. Sandrik, and H. Lopez, “Statistics of Speckle in Ultrasound B-Scans,” *IEEE Trans. Sonics Ultrason.*, vol. 30, no. 3, pp. 156–163, 1983, doi: 10.1109/T-SU.1983.31404.

[12] R. L. T. Bevan, J. Zhang, N. Budyn, A. J. Croxford, and P. D. Wilcox, “Experimental Quantification of Noise in Linear Ultrasonic Imaging,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 66, no. 1, pp. 79–90, 2019, doi: 10.1109/TUFFC.2018.2874720.

[13] N. Budyn, R. L. T. Bevan, J. Zhang, A. J. Croxford, and P. D. Wilcox, “A Model for Multiview Ultrasonic Array Inspection of Small Two-Dimensional Defects,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 66, no. 6, pp. 1129–1139, 2019, doi: 10.1109/TUFFC.2019.2909988.

[14] N. Brierley, T. Tippetts, and P. Cawley, “Data fusion for automated non-destructive inspection,” *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 470, no. 2167, 2014, doi: 10.1098/rspa.2014.0167.

[15] T. M. Loughin, “A systematic comparison of methods for combining p-values from independent tests,” *Comput. Stat. Data Anal.*, vol. 47, no. 3, pp. 467–485, 2004, doi: 10.1016/j.csda.2003.11.020.

[16] N. A. Heard and P. Rubin-Delanchy, “Choosing between methods of combining p-values,” *Biometrika*, vol. 105, no. 1, pp. 239–246, 2018, doi: 10.1093/biomet/asx076.

[17] L. H. C. Tippett, *The Methods of Statistics - An introduction mainly for workers in the biological sciences*, Third edit. Williams and Norgate Ltd. Great Russell Street, London, 1931.

[18] R. Fisher, *Statistical methods for research workers*. Oliver & Boyd, Edinburgh, 1925.

[19] H. A. Bloxham, A. Velichko, and P. D. Wilcox, “Combining Simulated and Experimental Data to Simulate Ultrasonic Array Data From Defects in Materials With High Structural Noise,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 63, no. 12, pp. 2198–2206, 2016, doi: 10.1109/TUFFC.2016.2614492.

[20] H. A. Bloxham, A. Velichko, and P. D. Wilcox, “Establishing the Limits of Validity of the Superposition of

Experimental and Analytical Ultrasonic Responses for Simulating Imaging Data," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 66, no. 1, pp. 101–108, 2019, doi: 10.1109/TUFFC.2018.2875781.

[21] S. O. Rice, "Mathematical Analysis of Random Noise," *Bell Syst. Tech. J.*, vol. 23, no. 3, pp. 282–332, 1944, doi: 10.1002/j.1538-7305.1944.tb00874.x.

[22] D. O. North, "An Analysis of the Factors which Determine Signal/Noise Discrimination in Pulsed-Carrier Systems," *Proc. IEEE*, vol. 51, no. 7, pp. 1016–1027, 1963, doi: 10.1109/PROC.1963.2383.

[23] R. D. Hippenstiel, *Detection theory : applications and digital signal processing*. CRC Press, 2002.

[24] P. Huthwaite, "Accelerated finite element elastodynamic simulations using the GPU," *J. Comput. Phys.*, vol. 257, pp. 687–707, 2014, doi: 10.1016/j.jcp.2013.10.017.