Robust principal component analysis of ultrasonic sectorial scans for defect detection in weld inspection

B. Cassels, L.-K. Shark, S. J. Mein, A. Nixon, T. Barber, et al.

B. Cassels, L.-K. Shark, S. J. Mein, A. Nixon, T. Barber, R. Turner, "Robust principal component analysis of ultrasonic sectorial scans for defect detection in weld inspection," Proc. SPIE 11059, Multimodal Sensing: Technologies and Applications, 110590E (21 June 2019); doi: 10.1117/12.2527622

SPIE.

Event: SPIE Optical Metrology, 2019, Munich, Germany

Robust principal component analysis of ultrasonic sectorial scans for defect detection in weld inspection

B. Cassels[a], L-K. Shark[a], S.J. Mein[a], A. Nixon[b], T. Barber[b], and R. Turner[b]

[a] University of Central Lancashire, Preston, UK;
[b] BAE SYSTEMS Maritime, Barrow-in-Furness, UK.

ABSTRACT

A single weld defect in a safety-critical engineering structure has the potential to incur high monetary costs, damage to the environment or loss of human life. This makes comprehensive non-destructive internal and external inspection of these welds essential. For non-destructive internal inspection the ultrasonic phased array supports a number of methods for producing a cross sectional image at a fixed location. Full coverage of the weld requires a sequence of images to be taken along the full length, each image at a unique incremental step. If the weld has a geometrically regular structure, such as that corresponding to a long linear section or the circumference of a pipe, automation becomes possible and data is now provided for post processing and auditing. Particularly in a production process this may provide many thousands of images a day, all of which must be manually examined by a qualified inspector. Presented in this paper is an approach for rapid identification of anomalies in sequences of ultrasonic sector images taken at equally spaced index points. The proposed method is based on robust principal component analysis (PCA). An assumption is that most sectors are anomaly free and have a statistically similar geometrical structure. Unsupervised multivariate statistical analysis is now performed to yield an initial low dimensional principal subspace representing the variation of the common weld background. Using the Mahalanobis distance outliers, observations with extreme variations and likely to correspond to sector scans containing anomalies, are removed from the reference set. This ensures a robust PCA-based reference model for weld background, against which a sectorial scan is identified as defect free or not. Using a comprehensive set of sector scan data acquired from test blocks, containing different types and sizes of weld defects at different locations and orientations, the paper concludes that PCA has potential for anomaly detection in this context. Although trimming improves the accuracy of the system eigenvectors, it is shown that greater accuracy of the low rank subspace is possible through principal component pursuit (PCP). This is evident by an almost 100% anomaly detection rate with a false alarm rate of well below 10%.

Keywords: Phased array, A-scan, Sector scan, ultrasonics, Anomaly detection, PCA, PCP

1. INTRODUCTION

Traditionally ultrasonic inspection is performed manually by a qualified inspector using a single element probe. Through a sequence of adaptive closed loop manual operations the inspector, through skill and experience, is able to detect and reliably sentence any anomaly within the weld. Limitations of this approach are that it is time consuming, full coverage of the weld is at the discretion of the inspector and there is no record of the ultrasonic data for future auditing purposes.

By the mid 1990's inexpensive and portable multichannel phased array (PA) systems were gaining in popularity. A significant advantage of a PA over a single element probe is that pulsing the elements in a timed sequence[1,2] allows the ultrasonic beam to be steered and focussed (figure 2). A common set up for weld inspection is illustrated in figure 3. Here the beam is steered (θ_s) through a discrete range of angles. At each angle the beam's focus is the weld's vertical centre line. Reflections from each angle (an A-scan) are recorded. After processing, this matrix of data is presented as a sectorial scan (figure 4.d) representing a cross sectional image of the weld. Where the weld geometry is consistent such as that around a pipe or a long linear butt weld, a robotic

Further author information: (Send correspondence to L-K.S)
B.C.: E-mail: bcassels1@uclan.ac.uk
L.K.S.: E-mail: lshark@uclan.ac.uk, Telephone: 044 1772 893xxx

Multimodal Sensing: Technologies and Applications, edited by Ettore Stella, Shahriar Negahdaripour, Dariusz Ceglarek, Christian Möller, Proc. of SPIE Vol. 11059 110590E · © 2019 SPIE · CCC code: 0277-786X/19/$21 · doi: 10.1117/12.2527622

manipulator may be used to incrementally move the PA along the weld in fixed length steps (indexing) before initiating the next data acquisition. Automated Ultrasonic Testing (AUT) in this manner provides a full record of the weld's condition at a point in time. In a production environment the approach potentially produces many thousands of images a day, each of which must be manually inspected to ensure it is anomaly free.

Whilst an inspector's experience can compensate for the negative effect of the time pressure and mental workload[3], other studies of human factors affecting the reliability of manual inspection[4,5], confirm that even in the hands of the most experienced inspectors, there remains some variability in the probability of anomaly detection and sentencing. The many thousands of cross sectional images requiring manual inspection only adds to inspection costs and demands placed on inspectors. In the worst case anomalies may be missed altogether.

To assist with the inspection of these vast quantities of data it is natural to consider machine vision and pattern recognition techniques. The earliest reports applicable to ultrasonic Non-Destructive Test (NDT) stem from the 1990's. Much of this early work tended towards the use of neural network and other artificial intelligence techniques to identify and classify individual faults[6,7]. Today this theme continues[8,9] but the indications are that research remains largely directed towards the classification and sizing of individual anomalies after their detection. Although of value in classifying anomalies, these techniques are not efficient in terms of initially identifying the areas for analysis. An assertion of this work is that in today's environment of automated data acquisition there is a need to rapidly locate potential anomalies. For industrial applications on the magnitude envisaged here it then becomes possible to apply the wealth of previous research to determine the type and size of faults identified. Even then, due to the critical nature of some structures, it is acknowledged that machine vision and pattern recognition techniques can only assist with this process. Final signing-off of a welded structure remains the judgement of a qualified professional.

This work represents an early approach at automatic anomaly detection in these vast quantities of data. No attempt is made to further classify any anomaly. For this investigation data is acquired from two carbon steel test blocks with induced defects. A two dimensional image of a weld is first created from the peak value in each reflected ultrasonic beam. Although thresholding the resulting image facilitates automatic identification of the larger faults, many anomalies are not detected. To improve the probability of detection (POD) the investigation considers a scheme based on principal component analysis (PCA). The underlying assumption is that anomalies appear as outliers to the normal data. Initially the training set consists of the entire data set. This is refined by ejecting observations containing significant variations. Trimming in this way leads to a more accurate description of the anomaly free subspace and, consequently, a more robust classification.

Section 2 provides an overview of the PA setup used to collect the data. Although test blocks are used the method of acquisition is, otherwise, that used in practice. For brevity details of equipment calibration and set up are excluded; however these are the same as those used in practice and are consistent for all data acquisitions of this type of weld. Section 3 outlines the procedure for creating 2D images and presents details of the methods of evaluation used through the remainder of the work. The results of thresholding provide a basis with which to evaluate improvements using PCA. Section 4 provides background information on the proposed scheme. Section 5 details an enhancement by principal component pursuit (PCP).

2. EXPERIMENTAL PROCEDURE

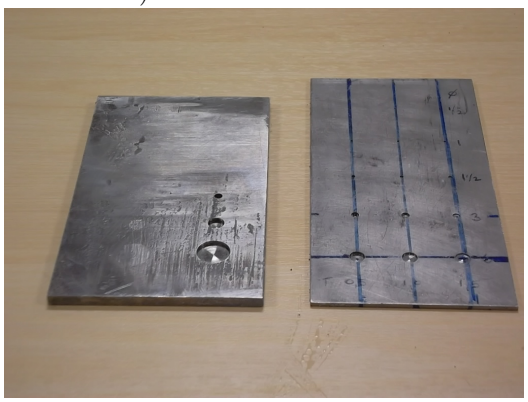
This section provides details of the test blocks to be used, the method of operating the PA and the format of the acquired data.

2.1 Test Pieces

To examine the performance of an ultrasonic test it is frequent to use test pieces with machined reflectors (for example, side drilled holes, flat bottomed holes (figure 1.a), notches and grooves). However it is suggested[10] that although these have uses in demonstrating the capability of a technique the results may be optimistic, and that artificial reflectors alone may lead to the development of an NDT technique that is hazardous when applied to structures containing real defects.

In an attempt to overcome these limitations welded test blocks, with induced defects representing cracks, lack of side wall fusion, pores and porosity, are available. Various techniques are used to implant faults. For example

a) Flat Bottomed Holes



b) Welded Test Block

Figure 1: Example test pieces - a) artifical reflectors, b) induced reflectors

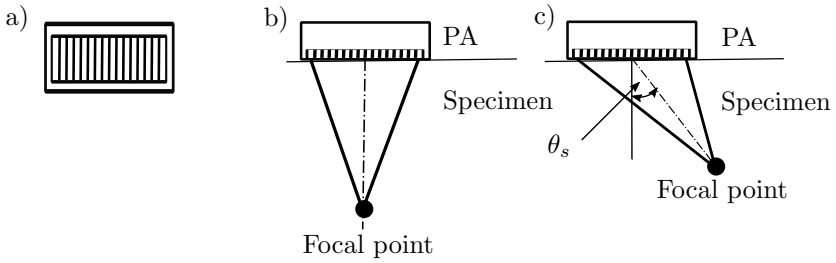


Figure 2: The phased array - a) footprint, b) focus only, c) focus and steering (θ_s)

an insert may be used to simulate lack of side wall fusion and a crack prone filler material may be used to induce a crack[11]. These give results that are more representative of anomalies found in a poor weld. Test data is from two welded test blocks (TB1 and TB2), similar to that illustrated in figure 1.b. Each contains a double-V butt weld. The material is of the same thickness (44 mm) and type of carbon steel.

2.2 The phased array

Figure 2.a illustrates the footprint of a linear phased array (PA) containing a row of rectangular piezo crystal elements separated by a constant gap and having an equal pitch. Exciting each element of the PA in a timed order enables the beam to be focussed (within the near field) and steered. In the diagrams (figure 2.b and c the dashed lines represent the effective beam path resulting from the application of a single focal law. Further information on PA operation is given Ditchburn and Ibrahim[1] and Drinkwater and Wilcox[2]. For this work it is only necessary to acknowledge that the PA can be operated in this manner and that the response to each focal law (termed an A-scan) is along a beam path represented by the dashed lines in the figures (2.b and c).

2.3 Physical set-up

Unlike figure 2 which provides a brief overview of the PAs capability, figure 3 is more representative of the set up used here. Significantly rather than being oriented above, and normal, to the weld the PA is mounted on an angled wedge and is offset from the weld's vertical centre line. As well as directing each beam into the weld refraction at the wedge to specimen interface results in a change in the dominant mode of wave propagation. Within the wedge propagation is dominantly longitudinal. Within the specimen, and through a process of mode conversion, the dominant propagation is by means of a shear (transverse) wave. The shear wave has a lower velocity and hence a smaller wavelength than the corresponding longitudinal wave. Along with details such as the respective sound velocities in the wedge and specimen, probe frequency and physical dimensions it is possible to write a set of focal laws for all beam angles within the specimen.

In figure 4.a a probe is dragged along the length of the test block in the direction of PA1. In practice it is sometimes desirable to scan the weld twice, once from each side (e.g. PA1 and PA2). This improves the

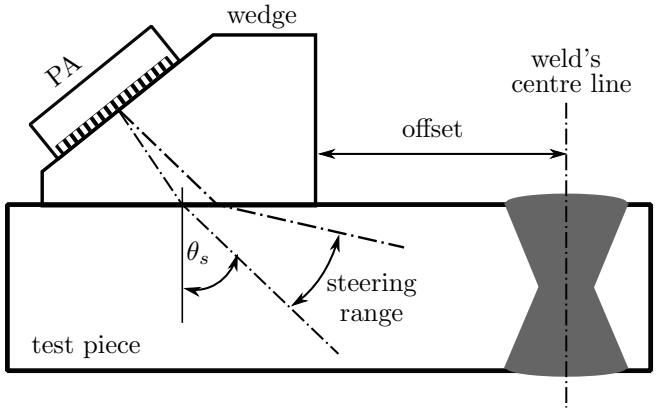


Figure 3: Overview of the physical set up (θ_s - steering angle)

probability of detecting a fault but, of course, increases acquisition time and data storage requirements. For these test blocks each weld is inspected four times. Each data set is denoted PA1, PA2, PA3 and PA4. In practice accessibility and time will likely prohibit all 4 scans, particularly if the weld is for two pipe sections. Here the goal is to detect all anomalies regardless of the particular PA direction.

2.4 Sector scans

For each index point the weld is scanned through a range of 40° to 75° in steps of 0.5° using 71 focal laws, figure 4.b. At each angle the focal point is the weld's centre line. the figure also indicates that each beam path through the weld is at a different angle. An alternative description of the beam paths is provided by figure 4.c. This more clearly illustrates that the full sector scan covers the weld at least twice. One consequence of this is that a single anomaly is also scanned at least twice. Depending on its size and location a single anomaly may, in fact, be scanned three times. Multiple instances of the same fault in a sector may by itself improve the POD. Reflection is greatest if the A-scan's angle is normal to the surface of the anomaly. Depending on its size and location reflection from one instance may be far stronger than from another. In an extreme case one instance may be indistinguishable from noise while the second instance is strongly pronounced.

An ultrasonic sector or S-scan, figure 4.d, is a collection of the resulting A-scans from a full set of focal laws at a single index position. In this particular case te figure illustrates two instances of a single anomaly, one instance being larger and stronger than the second. The figure also demonstrates that the recording of each A-scan does not start until a time determined by a gate. This removes unwanted front wall information and explains the curved start to the image. Similarly, because each A-scan is of the same length, the end of the sector scan is also curved.

The manually drawn overlay on the S-scan represents the unfolded scan paths and some care needs to be taken with interpretation of the y-axis on the figure. In particular it should be noted that the weld outline, representative of the first reflection, is actually inverted about the vertical axis (e.g. 4.c).

In terms of analysis this work primarily uses A-scans as a set of vectors each of length 1544 samples. A matrix of 71×1544 representing each sector. Here images such as figure 4.d are largely for visual information only. Details of image construction are, therefore, not described.

3. CREATION AND THRESHOLDING OF 2D IMAGES

The test data consists of a number of sector scans taken at equally spaced index points along the length of a weld and at a fixed offset (one data set for each PA direction). In the first instance converting this volumetric description to two dimensions facilitates anomaly detection. This conversion is either by taking the peak value of each A-scan (5.a) or the peak value of the same sample point in each A-scan (figure 5.b). The first case produces a 2D image (C-scan) from the viewpoint of the PA, the second (neglecting the curvature of the sample points) a viewpoint from above the weld. Either viewpoint aids the location of anomalies. As the PA viewpoint is of smaller dimensions this is, for this work, the preferred viewpoint.

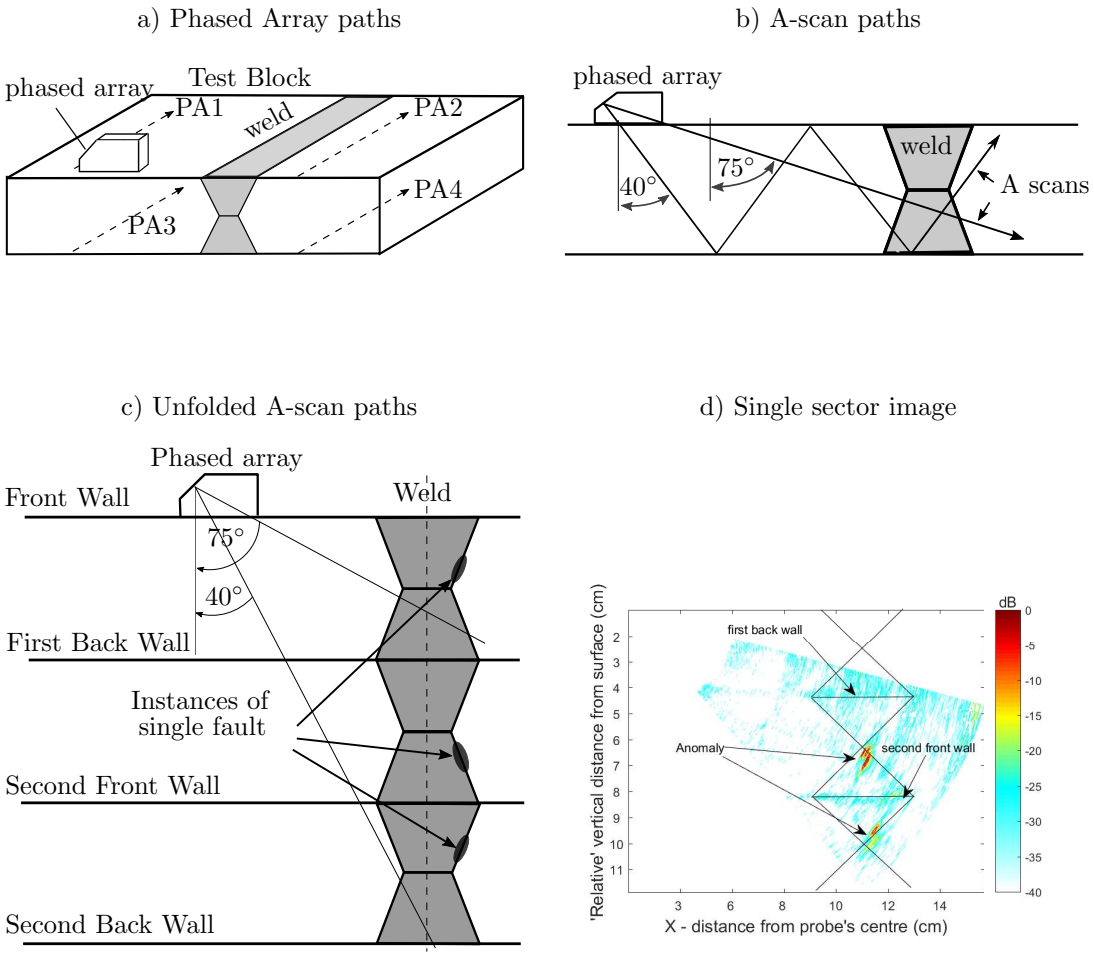


Figure 4: Overview of weld observation using a phased array

3.1 Thresholding and ground truth comparisons

One of the simplest methods of automatically identifying anomalous regions is to apply soft thresholding to each image. To recap there are two test blocks (TB1 and TB2) and each is scanned in 4 directions (PA1, PA2, PA3 and PA4). Figure 6 provides an illustration of the peak A-scan values for each of the 8 data sets and corresponding ground truth templates. In all cases the ground truth template is established by first using the data sheet provided by the test block manufacturer. This information is refined from a manual ultrasonic inspection of each test block by a qualified inspector. The templates are unlikely to be an exact representation of the respective test piece. They will, however, be very close and provide a relative measure of performance when comparing different methods of classification.

Chosen primarily for their popularity three methods of thresholding (Kittler Illingworth,[12] Otsu[13] and Maximum Entropy[14]) are attempted. After applying the respective threshold to each image a comparison is made with the appropriate ground truth. A quantitative comparison can be made on the basis of the confusion statistics (where TP = true positive, TN = true negative, FP = false positive and FN = false negative). The simplest method of comparison is based on a measure of accuracy (the percentage of pixels correctly identified as TP and TN):-

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

A primary problem of equation 1 when used as a single metric is that it suffers from the accuracy paradox[15]. This is a particular problem when there is an imbalance between the number of false positives and false negatives[16].

a) Peak A-scan sample values ('plan' view)

b) Peak A-scan values (PA view)

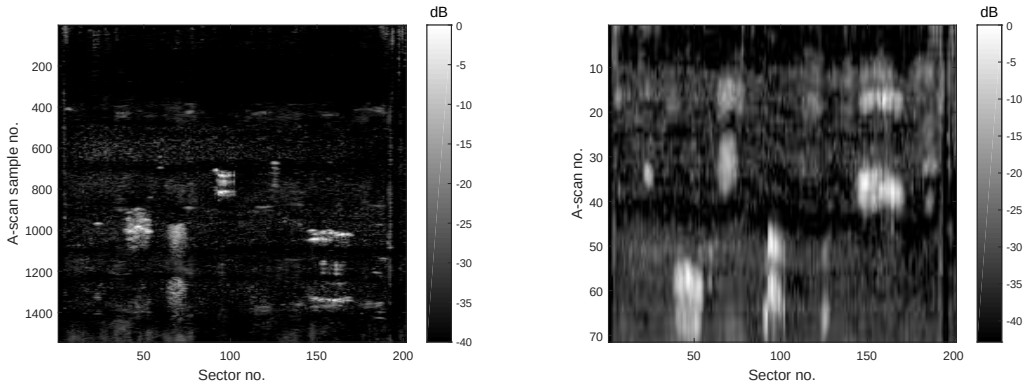


Figure 5: Example 2D images from the two viewpoints

Image | Ground truth | Image | Ground truth

a) TB1PA1 b) TB1PA1 c) TB1PA2 d) TB1PA2

e) TB1PA3 f) TB1PA3 g) TB1PA4 h) TB1PA4

i) TB2PA1 j) TB2PA1 k) TB2PA2 l) TB2PA2

m) TB2PA3 n) TB2PA3 o) TB2PA4 p) TB2PA4

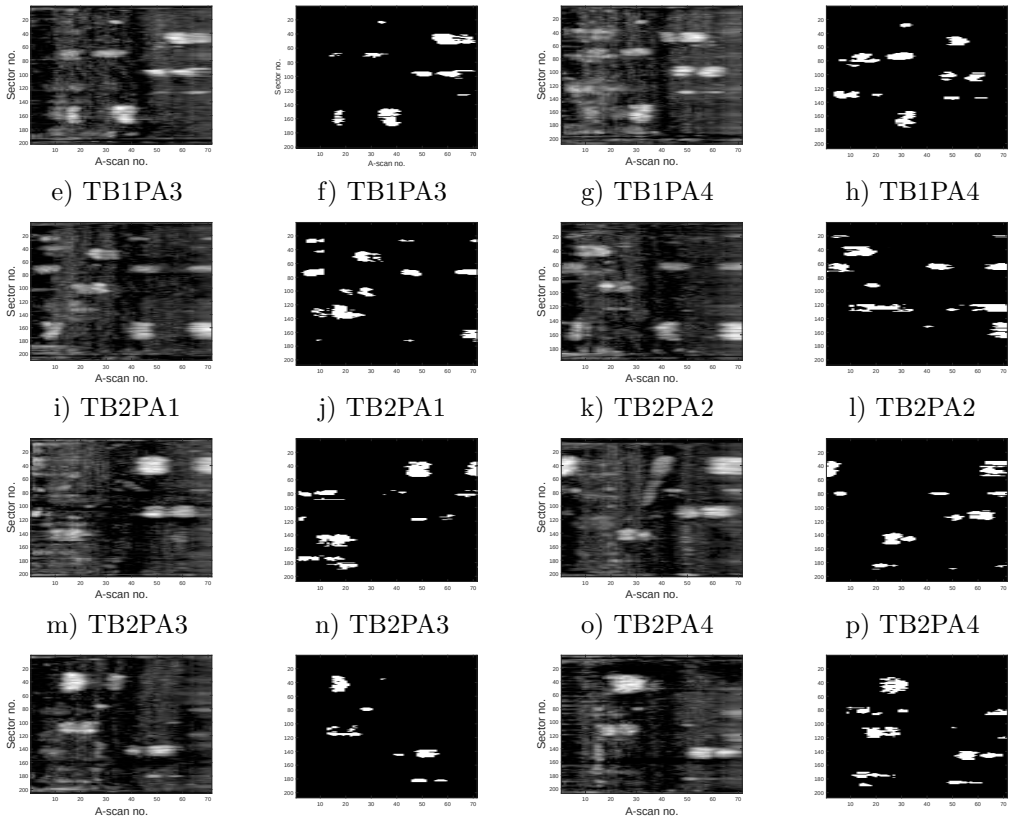


Figure 6: Columns 1 and 3 peak A-scan values (db scale 0 to -40 dB), columns 2 and 4 ground truth masks

Although low accuracy indicates a poor classification, high accuracy, by itself, does not necessarily indicate good classification.

For weld inspection, identification of all anomalies is of paramount importance. This can be at the expense of false positives provided these numbers are not so large that users will lose confidence in the technique. In this respect two measures, sensitivity and specificity, are needed.

If the classification detects all true positives then there will be no false negatives. Recall or sensitivity defined as:-

$$sensitivity = \frac{TP}{TP + FN} \tag{2}$$

is a measure of how often the classifier predicts a true positive. Ideally this will be 1.

However, if the threshold is too low, many false positives will lead to many false alarms. Ideally, for detection of all true negatives, there will be no false positives. Fall out or specificity defined as:-

$$specificity = \frac{TN}{TN + FP} \tag{3}$$

is a measure of how often the classifier predicts a true negative. Ideally this will also be 1.

The ROC curve[17] is a two-dimensional chart which combines the two metrics. This is done by plotting the fall-out or false positive rate (1-specificity) against sensitivity. An early adaptation of this as a technique to evaluate different classification schemes was by Spackman[18]. The approach is seeing increasing use[19]. It is unaffected by imbalances in the confusion statistics[20] and has two properties of value, discussed with reference to figure 7.

a) ROC example 1 b) ROC example 2

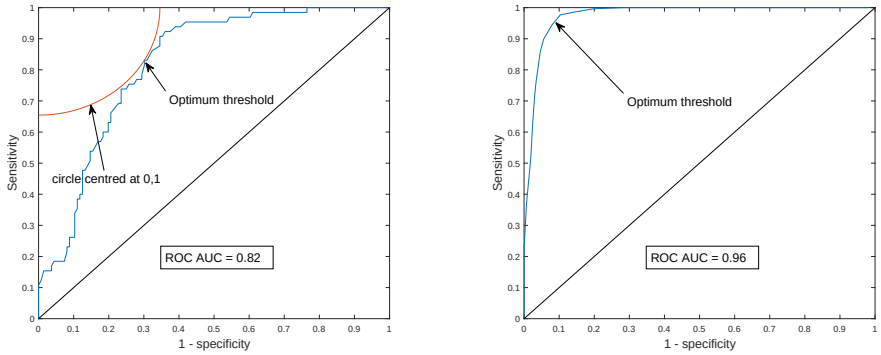


Figure 7: Example ROC curves

Firstly, the Area Under the Curve (AUC) provides a measure of the effectiveness of the classification. Secondly, as used here, the ROC curve provides an 'optimal' trade-off between sensitivity and specificity, Greiner et al.[21]. Where the 'optimal' threshold is defined as the point closest to the curve's top left hand corner.

Consider the two examples, figure 7. A visual inspection indicates example 2 to be a better classifier than example 1. This is confirmed by the respective AUCs of 0.96 and 0.82. For an ideal classifier the curve will touch the top left corner (0,1) with an AUC = 1. The line from (0,0) to (1,1) represents an AUC of 0.5; at this level a classifier has no use.

Table 1 lists a set of statistics obtained from thresholding each set of PA results using the three thresholding methods. The average values (last row) indicates the KI method to be the superior in this case. However the sensitivity figures for TB1PA2 and TB2PA4 are poor. The last 3 columns (ROC statistics) provide the result of a ROC analysis. This is by sweeping the threshold value through a range of values. Each point on the ROC curve corresponds to the respective sensitivity and specificity value of each threshold. In each case the AUC indicates good classification to be possible. A comparison of the ROC thresholds indicates a closer correspondence to

Table 1: Comparison of thresholding methods

	Kittler Illingworth		Otsu		Maximum Entropy		ROC statistics		
Test Piece	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.	AUC	Sens.	Spec.
TB1PA1	1.0	0.95	0.75	0.99	0.67	0.99	0.99	1.0	0.95
TB1PA2	0.32	0.99	0.24	0.99	0.71	0.97	0.95	0.98	0.90
TB1PA3	1.0	0.95	0.50	0.99	0.27	0.99	0.95	0.99	0.95
TB1PA4	0.93	0.96	0.19	0.99	0.2	0.99	0.93	0.98	0.93
TB2PA1	0.97	0.96	0.43	0.99	0.44	1.0	0.95	0.97	0.97
TB2PA2	0.91	0.96	0.43	0.99	0.7	0.98	0.93	0.98	0.92
TB2PA3	0.96	0.96	0.49	0.99	0.61	0.99	0.93	0.96	0.97
TB2PA4	0.79	0.96	0.79	0.96	0.45	0.99	0.89	0.97	0.90
Ave's	0.86	0.94	0.48	0.99	0.51	0.99	0.94	0.98	0.94

those obtained using the KI method. In particular there is a 12% improvement in average sensitivity with no reduction in specificity. The ROC curve indicates that an improved threshold may be possible, but provides no indication of how to achieve this value without a ground truth.

3.2 Limitations of thresholding

In these cases table 1 indicates that the KI threshold produces the higher performance. However in instances, such as TB1PA2, the sensitivity is particularly poor suggesting that under some circumstance the method may not be reliable. In fact the table also indicates some considerable variations in sensitivity for the Otsu and maximum entropy methods. An additional problem associated with thresholding is that a value will often be found. All of the samples available have a relatively dense distribution of of anomalies. Under normal circumstances, in practice, there is a high probability that large lengths of weld will be free from anomalies. In this case applying an automatic threshold is likely to indicate a high alarm rate.

4. MULTIVARIATE STATISTICAL ANALYSIS

For a data sets with a single variable there are likely to be some observations that do not belong to the pattern produced by others. If a normal distribution is assumed, then an observation with a standardised value more than, say, 3 standard deviations away from the distributions mean, might be considered an anomaly. An alterative to thresholding is to build a model of normal data and detect any deviation from this in the measured data[22]. There are in fact many applications where the classifier has to detect only whether an observation belongs to the data on which it was trained. In the literature these systems are often referred to as fault or novelty detection systems. A complication of this case, as with many anomaly detection systems, is that the data is high dimensional. For further exploration some simplification becomes highly desirable. The following describes a method based on principal component analysis[23] (PCA).

4.1 PCA

Rather than making a distributional assumption about the data PCA primarily explains the variance-covariance structure of a set of variables by a smaller number of variables (principal components) that are linear combinations of the original, mean centred, variables. The principal components are obtained from a singular value decomposition (SVD)[24] of the covariance or correlation matrix of the original, mean centred, variables.

Let the original data $\mathbf{X}$ represent an n (observations)$\times p$ (variables) matrix (X_1, X_2, ... X_p) and $\mathbf{S}$ be a p$\times$p covariance matrix of $\mathbf{X}$. Eigen analysis of the covariance matrix produces a set of eigenvalue/eigenvector pairs $(\lambda_1, e_1), (\lambda_1, e_1), \cdots, (\lambda_p, e_p)$ and the i_{th} principal component is now:-

$$y_i = \mathbf{e_i^T}(\mathbf{X} - \overline{\mathbf{X}}) = e_{i1}(x_1 - \overline{x}_1) + e_{i2}(x_2 - \overline{x}_2) + ... + e_{ip}(x_p - \overline{x}_p), \quad i = 1, 2...p. \tag{4}$$

where:

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$

$\mathbf{e}_i^T = (e_{i1}, e_{i2}, \cdots, e_{ip})$ is the i_{th} eigenvector

$\mathbf{X}^T = (x_1, x_2, \cdots, x_p)$ is any observation vector of variables $X_1, X_2, \cdots, X_p$,

and $\overline{\mathbf{X}}^T = (\overline{x}_1, \overline{x}_2, \cdots, \overline{x}_p)$ is the sample mean vector of variables $X_1, X_2, \cdots, X_P$.

The principal components have three important properties 1) they are uncorrelated, 2) they are ordered in terms of descending variance, 3) the total variation in all principal components is equal to the total variance in the original variables.

4.2 Anomaly detection

For anomaly detection using PCA the sum of squares of the principal component scores (equation 4):-

$$\sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \cdots + \frac{y_p^2}{\lambda_p} \tag{5}$$

is the same as Mahalanobis distance[25] of observation $\mathbf{X}$ from the mean of the sample.

Under the previous conditions ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$) and assuming a large sample size the first q principal components have a chi-square distribution[24] with q degrees of freedom. Given a significance level α, an observation is an outlier if $\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i} > \chi^2_{(q,\alpha)}$.

4.3 The training set

It is important that the training set is an accurate estimate of anomaly free data. For PCA a 'robust fit' refers to the process of finding an estimation of the eigenvectors that would exist without contamination. Principal amongst methods of achieing this are the Maximum Covariance Determinant (MCD) and Minimum Volume Ellipsoid. The two have similarities; however a significant difference is that for a predefined number of points MCD looks for a subset of data points whose covariance matrix has the smallest determinant, whilst MVE searches for an ellipsoid with smallest volume. These and similar approaches have roots in earlier works by Rousseeuw[26,27] and Rousseeuw and Leroy[28]. For this work the primary drawback of the MVE and MCD procedures is the computational effort required. Recent publications evaluate procedures for fast-MCD and fast-MVE. However the applications remain limited to around 100 variables[29,30].

Multivariate trimming is a simpler method of obtaining a robust training set and suitable for unsupervised training. This begins by taking the entire data set and determining the Mahalanobis distance for each observation. Assuming that the number of normal observations is much larger than the number of anomalies the observations with large Mahalanobis distances are ejected from the training set before obtaining new, trimmed, estimators for $\overline{\mathbf{X}}$ and $\mathbf{S}$. In both of these respects the data available from the test blocks has some limitations. In particular the number of observations is low in terms of the dimensionality and there is a relatively high density of sectors containing anomalies.

An example of these limitations is provided by the first investigation where the training set consists of observations for each sector. This is similar to the approach taken in many image recognition problems[31,32] where it is not uncommon to concatenate all rows of an image into a single vector. Here a sector contains 71 A-scans, each with 1544 samples. The overall vector length, and number of dimensions (p), is 109624. This is far greater than the number of observations (around 1600). In addition the resulting covariance matrix is 109624×109624 ($p \times p$). Although determining the eigenvectors of such a large matrix is not intractable on a modern computer it consumes considerable processing power and time.

Taking note of these limitations the first investigation is to create a training set using each sector as an observation. As the maximum number of eigenvalues is limited by the number of observations ($n-1$), all others having a value of zero, a computationally easier option proposed by Turk and Pentland[33] is to solve the transpose of the data matrix (leading to a $n \times n$ covariance matrix) to produce training set's n, none zero, eigenvectors.

4.4 Full sector training sets

Individual observations represent a single full sector. Each mean centred observation is now projected against the training set's eigenvectors (equation 4). However only the first few eigenvectors (q, where $q < p$), those with the largest eigenvalues, are used. These correspond to the first q principal components. Establishing how many PCs to retain (the so called 'stopping rule') is a matter of some discussion[34]. Graphical techniques such as the 'elbow' or 'hockey stick' methods[35] are subjective, can have a tendency to underestimate the dimension of the data[35] and are not suitable for automation. Other alternatives[36] include a method due to Kaiser[37] where the number of eigenvalues greater than the average of all (p) eigenvalues is taken as the number of principal components (q) to use. It is the method used here.

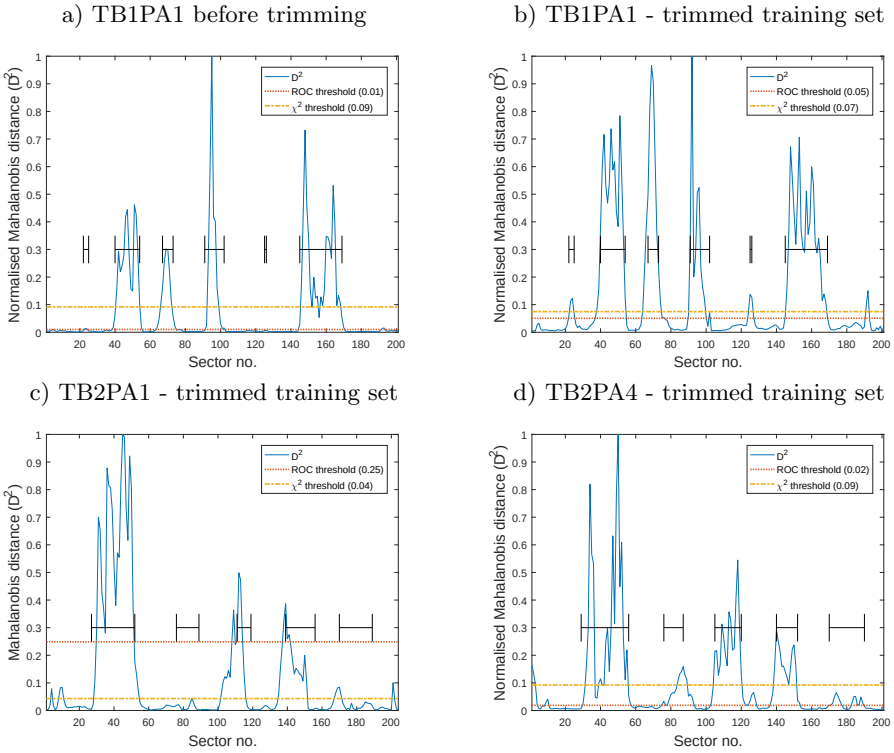


Figure 8: Examples of training set trimming - Full sector projections

Figure 8.a and b illustrates the result of projecting TB1PA1 against the untrimmed and trimmed training sets; each superimposed 'H' shape covers a range of sectors containing an anomaly from the ground truth. The two dashed horizontal lines respectively represent the $\chi^2_{(q,0.95)}$ limit and the ROC threshold. Any sector with a Mahalanobis distance greater than the $\chi^2_{(q,0.95)}$ limit is automatically flagged as an anomaly. The ROC threshold is derived by scanning through a range of thresholds to determine a theoretically optimum value (section 3.1). In this case a visual comparison of the two indicates the improvements due to trimming. In particular the two small anomalies masked out by the untrimmed training set are visible after trimming. Table 2 provides a set of sensitivity and specificity figures for each test piece. In all cases the sensitivity figures increase after trimming but overall the results do not suggest the approach to be of practical use. In fact the ROC statistics are not encouraging and perhaps more importantly the table does not indicate that, for the full set of test pieces, there is no indication of 6 anomalous instances, although this is an improvement over 14 missed instances for the untrimmed training set.

These results are disappointing but indicate some ability of the technique for anomaly detection in this context. The primary difficulty is in the high dimensionality of the data and the high density of anomalies in the test

Table 2: Discrimination before and after trimming

Test piece	Untrimmed (39 PCs)			Trimmed (39 PCs)			ROC trimmed			
	Acc.(%)	Sens.	Spec.	Acc.(%)	Sens.	Spec.	Acc.(%)	AUC	Sens.	Spec.
TB1PA1	91.0	0.74	0.99	92.5	0.89	0.94	93.0	0.98	0.98	0.90
TB1PA2	76.4	0.56	0.92	71.1	0.70	0.72	75.5	0.79	0.53	0.93
TB1PA3	76.1	0.47	0.96	78.9	0.64	0.90	80.4	0.88	0.89	0.74
TB1PA4	81.6	0.53	0.97	81.6	0.72	0.87	82.1	0.88	0.74	0.87
TB2PA1	69.6	0.34	0.96	69.6	0.53	0.82	70.6	0.74	0.33	0.98
TB2PA2	80.9	0.53	0.99	83.8	0.77	0.88	85.8	0.88	0.74	0.94
TB2PA3	82.4	0.49	0.99	84.8	0.72	0.91	82.4	0.91	0.87	0.80
TB2PA4	70.8	0.34	0.99	78.2	0.59	0.93	81.6	0.88	0.92	0.73
Averages	83.5	0.52	0.98	80.1	0.69	0.87	81.4	0.87	0.75	0.86

pieces. Nothing can be done with respect to the high density of anomalies. However a training set with a smaller dimensionality can be achieved by considering an alternative viewpoint.

4.5 Constant angle A-scans

Rather than convert an entire sector into a single observation an alternative is to use a set of constant angle A-scans as the training set. Under this scenario the number of dimensions reduces to the number of samples in each A-scan. Figure 9 illustrates this arrangement. For the test pieces under consideration the number of dimensions reduces to 1544; a considerable reduction on the full sector approach. A complication is that a number of training sets are now required (one for each angle) and tests need to be repeated at every angle rather than the sector as a whole. Another alternative is to consider a group of constant sample numbers as the training set. Although this reduces the dimensions to the number of A-scans (71) this requires 1544 training sets. In the following the first approach, that of constant angle A-scans, only, is considered due to its relatively smaller number of training sets.

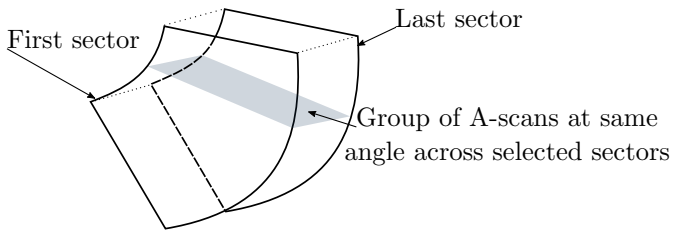


Figure 9: Training set as A-scans at constant angle

Observations at a specific A-scan angle are now projected against the respective training set. For each test piece a sweep through all sectors of a specific A-scan observation against its respective training set leads to a profile of D^2 distances for all sectors. Repeating this for all A-scans produces a visual representation of the test piece as illustrated in figure 10.a and b. As the number of principal components, q, is different for each training set, so too is the $\chi^2_{(q,0.95)}$ threshold. Before producing the figure each threshold is subtracted. Consequently any point greater than 0 is considered a potential anomaly.

After trimming, the most notable visual effect is a reduction in the number of low level D^2 distances, probably false positives, breaking through the floor (0 level) of the plot. This is confirmed by comparing the respective binary masks (figures 10.c and d) with the ground truth mask (figures 10.e). Figure 10.f is the mask produced by using the 'optimum' ROC threshold for the trimmed data.

A more objective comparison is listed in table 3. In all cases the sensitivity is relatively high, the effect of trimming being to reduce the false alarm rate (higher specificities). Trimming also produces a more consistent set of sensitivities and specificities over all test pieces. Although the untrimmed data indicates some discrimination the trimmed case provides almost 100% sensitivity with a marked improvement in the false alarm rate.

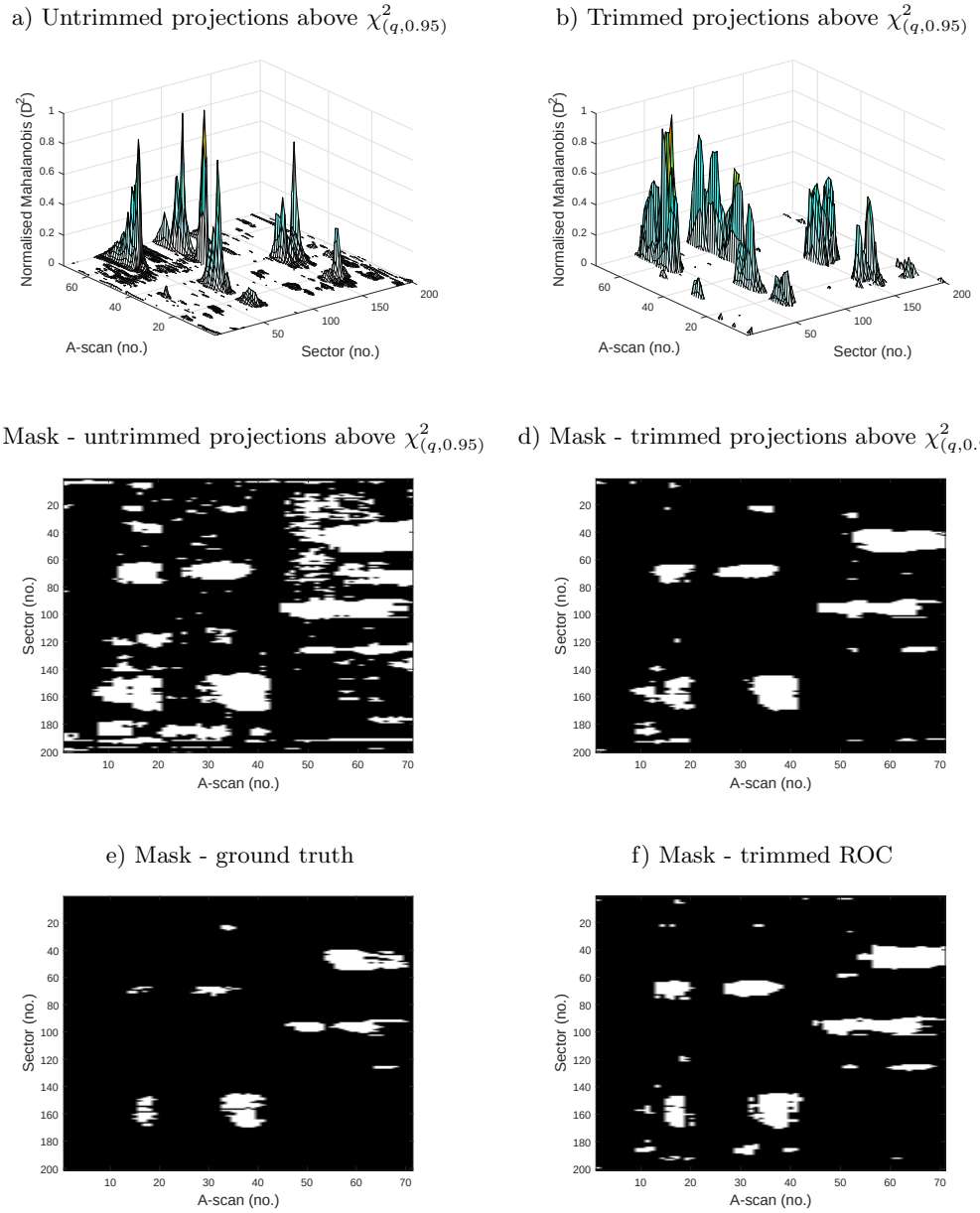


Figure 10: Examples (TB1PA1) of training set trimming - constant A-scan

4.6 Limitations

To improve robustness in the presence of outliers the previous section used trimming to improve the estimation of an anomaly free condition. This procedure alone suffers from a number of limitations, in particular:-

1. Outlier removal is through a single iteration. Greater accuracy may be possible by repeating the procedure.
2. Only the major principal components ($q \leq p$) are used, no consideration has been given to the minor components. These may be inconsistent with the co-variance structure[30,25] making their ejection desirable.
3. Outlier detection is based only on a single distance measure (the Mahalanobis distance) and using the $\chi^2_{(pc,0.95)}$ limit.

Table 3: Constant A-scan look-up (untrimmed, trimmed and ROC) with all reference data

Test Piece	Untrimmed			Trimmed			ROC			
	Acc.(%)	sens.	spec.	Acc.(%)	Sens.	Spec	AUC	Acc.(%)	Sens.	spec
TB1PA1	83.9	0.99	0.83	96.1	0.99	0.85	0.99	95.0	0.96	0.94
TB1PA2	69.5	0.99	0.68	81.5	0.99	0.93	0.98	95.3	0.93	0.95
TB1PA3	84.7	0.99	0.84	89.6	1.0	0.90	0.99	95.1	0.96	0.95
TB1PA4	86.0	0.99	0.86	90.3	0.99	0.90	0.99	93.6	0.94	0.93
TB2PA1	81.3	0.92	0.81	87.9	0.99	0.88	0.95	95.9	0.89	0.96
TB2PA2	81.3	0.99	0.80	87.8	0.99	0.88	0.99	85.9	0.95	0.96
TB2PA3	89.0	0.94	0.89	91.8	0.98	0.92	0.97	94.4	0.91	0.94
TB2PA4	87.4	0.91	0.97	92.4	1.0	0.92	0.95	93.7	0.88	0.93
Averages	82.8	0.97	0.83	89.0	0.99	0.90	0.98	93.6	0.93	0.95

4. The number of PCs to use is that specified by the Kaiser rule.

Points 3 and 4 are a matter of some judgement. The values given have not been investigated in detail and they are known not to give 'optimum' results. However small variations, particularly in the number of principal components does not produce significant change in results. At present the aim is to provide as much automation as possible, therefore, the $\chi^2_{(pc,0.95)}$ limit and the Kaiser stopping rule are, for the time being maintained.

Methods of overcoming limitations raised in points 1 and 2 were alluded to in 4.3 and methods of improving robustness are discussed by Gnanadesikan and Kettenring[38]. More recent methods involve replacing the L2 by the L1 norm[39], a particular advantage being that the L1 norm is less sensitive to outliers. Driven by applications such as video surveillance, which may require a separation of static background from a moving foreground, the L1 norm has a further property of producing a sparse solution[40]. Briefly a matrix is sparse if it contains only a few entries that are non-zero. For a given matrix, sparsity is a measure of the number of non-zero elements divided by the total number of elements; the number of non-zeros is termed the L0 norm. A matrix X is now decomposed into low rank (L) and simpler sparse (S) components:-

$$X = L + S \tag{6}$$

If L and S represent accurate representations of the static (background) and dynamic (foreground) components then matrix L can be used as the training set in conventional PCA. A comprehensive review of many methods for solving this Robust PCA (RPCA) problem, where the data matrix is decomposed in this way is given by Bouwmans et al.[41]. One of their conclusions is that although, to date, no algorithm has emerged that is able to simultaneously address all of the key challenges of real video data processing, principal component pursuit (PCP), outlined by Candès et al.[42], provides the most representative formulation for the problem. A further strong motivation for using PCP is the availability of an implementation (through Stanford University) of the freely available Template for First Order Conic Solvers (TFOCS)[43] software. TFOCS provides a set of Matlab building blocks that can be used to solve a variety of convex models. However, for this specific application it was found more convenient to implement PCP in a simpler single (Matlab) function. Once again the Matlab function is available in open source software[44].

5. OVERVIEW OF PCP

Before presenting results based on RPCA an outline of the PCP algorithm is used. The objective is to split the matrix into a low rank and sparse components. From seminal works[42,45] recovery of a low rank matrix is obtained as follows:-

$$\begin{aligned} &\text{minimise} \quad \|L\|_* + \lambda\|S\|_1 \\ &\text{subject} \quad \text{to} \quad X = L + S \end{aligned} \tag{7}$$

Where $\|L\|_* = \Sigma_{i=1}^{r}\sigma_i(L)$ is the nuclear norm of L (the sum of singular values).

and $\|S\|_1 = \Sigma_{i,j}^{r} S_{i,j}$ is the L1 norm of the matrix S, thought of as a vector.

The parameter λ controls the weight to put on S relative to L. It is not overly sensitive. It can be chosen by a few iterations or by:-

$$\lambda = \frac{1}{\sqrt{min(r,c)}} \tag{8}$$

where r and c are the number of rows and columns respectively in the data matrix X

5.1 Robust PCA using principal component pursuit

For data sets of the size used here PCP is computationally intensive and time consuming. Results are therefore limited to the constant A-scan data sets. For this orientation PCP is used to create 71 low rank matrices from each corresponding group of original observations. These are then used with conventional PCA as previously described. With the default value of λ, equation 8, results were disappointing with no evidence of anomaly detection.

In the original work of Candès[42] the focus is on recovering a true low rank L and sparse S. A study by Paffenroth et al.[46] suggests that when used for anomaly detection the Candès approach, with the default value of λ is unlikely to be a suitable value for downstream anomaly detection. This is in line with the findings here. It is further suggested and demonstrated[46], that for anomaly detection λ is better viewed as a parameter to be estimated from the training data and tuned to the particular application. However this suggestion[46] was for a single application with access to far more data than is available here. In this case the data is limited in quantity and in practice there will be many different types of ultrasonic inspection, each representative of a different application. At present the interest here is to first investigate PCP as a method of providing an accurate sub-space of the particular data set. This is most simply done by repeatedly obtaining an L for different λs. Using the test vectors and a range of thresholds, a projection of each test set against L's sub-space produces a single set of confusion statistics for all 8 test pieces.

After a number of iterations λ was set to $10\times\lambda_{default}$ (where $\lambda_{default}$ is specified by equation 8).

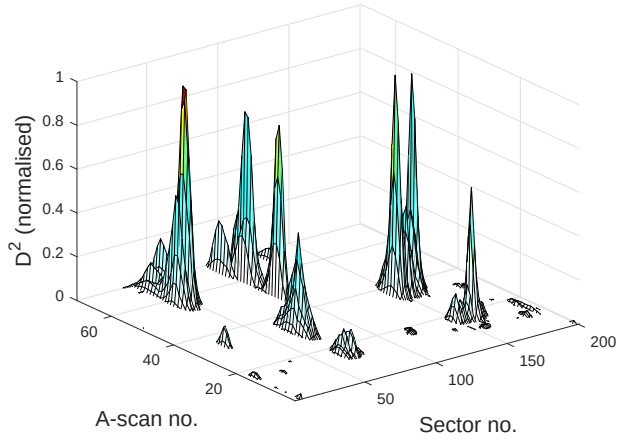
5.2 PCP results

Figure 11.a illustrates the result of a sector and A-scan look-up for TB1PA1. The respective confidence level is subtracted from each projection so that any value greater than zero indicates a potential anomaly. Figure 11.b is the corresponding mask for projections exceeding the $\chi^2_{(pc,0.95)}$ limit. Visually there is a remarkable similarity between the $\chi^2_{(pc,0.95)}$ mask and the ground truth. This is confirmed by table 4. In addition to sensitivity values of 1 in almost all cases the specificities average 0.95. ROC analysis indicates that a near perfect classification is now possible.

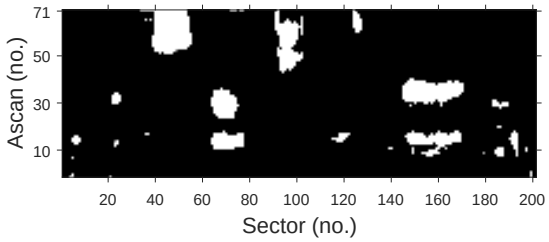
Table 4: Classification trimmed (from table 3) vs. PCP

Test Piece	Trimming $\chi^2_{(pc,0.95)}$			PCP $\chi^2_{(pc,0.95)}$			ROC			
	Acc.(%)	Sens.	Spec.	Acc.(%)	Sens.	Spec	AUC	Acc(%)	Sens.	Spec.
TB1PA1	83.9	1.00	0.83	95.7	1.00	0.96	0.99	98.5	0.99	0.99
TB1PA2	69.5	0.99	0.93	91.2	1.00	0.91	0.99	97.6	0.99	0.98
TB1PA3	84.7	0.99	0.84	94.3	1.00	0.94	0.99	98.4	0.99	0.98
TB1PA4	86.0	0.99	0.86	94.2	1.00	0.94	0.99	98.2	0.99	0.98
TB2PA1	81.3	0.92	0.81	94.4	1.00	0.94	0.99	98.0	0.99	0.98
TB2PA2	81.3	0.99	0.80	93.9	0.99	0.94	0.99	98.3	0.99	0.98
TB2PA3	89.0	0.94	0.90	96.5	1.00	0.96	0.99	99.2	0.99	0.99
TB2PA4	87.4	0.91	0.87	95.8	1.00	0.96	1.0	99.3	0.99	0.99
Averages	82.9	0.97	0.86	94.5	1.00	0.95	0.99	98.5	0.99	0.98

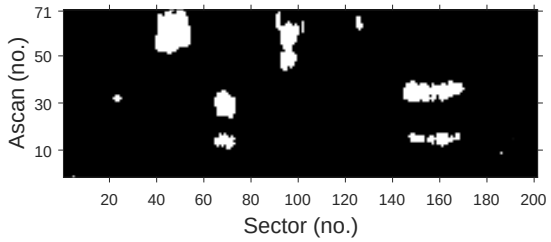
a) Normalised D^2 projections above confidence limit (PCP)



b) Mask - D^2 projections above $\chi^2_{(pc,0.95)}$



c) Mask - D^2 projections above ROC threshold



d) Mask - ground truth

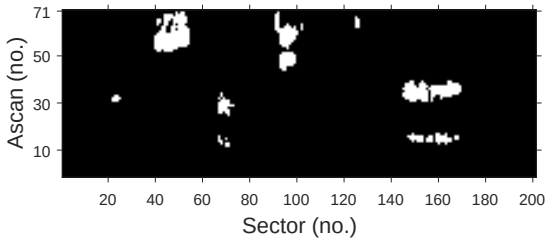


Figure 11: TB1PA1 projections ($\lambda = 10\times\lambda_{default}$)

6. COMMENTS AND FURTHER WORK

This work has demonstrated the potential capability of anomaly detection in welds using PCA. In place of a conventional C-scan, which creates a 2D image using peak A-scan values, the PCA approach is found to have the potential to highlight anomalies with a high POD and low false alarm rate.

Although data trimming, whereby observations representing potential outliers are ejected from the training set, is found to increase the POD a much improved subspace is obtained through PCP. Once obtained the low rank matrix may be used repeatedly for tests using the same material (dimensions and type) and set up.

Within the limitations of the available data sets there is confidence that results using the contact phased array are both repeatable and applicable to other situations. However the fact remains that the test pieces used are limited in size and a suggestion for further work must be a repetition of the technique on more extensive data sets and ultimately on production data.

REFERENCES

[1] R. J. Ditchburn and M. E. Ibrahim, "Ultrasonic Phased Arrays for the Inspection of Thick-Section Welds," tech. rep., Department-of-Defence, 04 2009.

[2] B. Drinkwater and P. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," *NDT & E International* **39**(7), pp. 525 – 541, 2006.

[3] M. Bertovic, B. Fahlbruch, C. Müller, J. Pitkänen, U. Ronneteg, M. Gaal, D. Kanzler, and U. E. D. Schombach, "Human factors approach to the acquisition and evaluation of NDT data," in *18^{th} World Conference on Nondestructive Testing, Durban, South Africa*, Apr. 2012.

[4] F. Fücsök, C. Müller, and M. Scharmach, "Human factors: The nde reliability of routine radiographic film evaluation," in *Proceedings of 15^{th} World Conference on Non Destructive Testing, Roma*, 10 2000.

[5] F. Fücsök, C. Müller, and M. Scharmach, "Reliability of routine radiographic film evaluation: An extended roc study of the human factor," in *Proceedings of 8th European Conference on Non Destructive Testing, Barcelona*, 6 2002.

[6] A. McNab and I. Dunlop, "Artificial intelligence techniques for the automated analysis of ultrasonic ndt data," *NDT & E International* **25**, pp. 6/1 – 6/8, 01 1991.

[7] A. Masnata and M. Sunseri, "Neural network classification of flaws detected by ultrasonic means," *NDT & E International* **29**, pp. 87–93, 04 1996.

[8] Q. Ri, R. Chang-Ming, and B. Yan-Ru, "The application of case-based reasoning to defect interpretation in non-destructive testing," in *Proceedings of 2004 International Conference on Machine Learning and Cybernetics(IEEE Cat. No.04EX826*, **4**, pp. 2296–3000, 08 2004.

[9] Z. Wang and X. Li, "Fuzzy variable method for characterizing flaw sizes," *Theoretical and Applied Fracture Mechanics* **63-64**, pp. 50 – 53, 2013.

[10] S. Crutzen, P. Lemaitre, and I. Iacono, "Realistic defects suitable for ISI (In Service Inspection) capability evaluation," in *NDE in the Nuclear and Pressure Vessel Industries. Poceedings, 14th International Conference*, 1996.

[11] M. Consonni, W. Fun, and C.Scneider, "Manufacturing of welded joints with realistic defects," in *50th Annual conference of the British Institue of Non-Destructive Testing*, Sept. 2011.

[12] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognition* **19**(1), pp. 41 – 47, 1986.

[13] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics* **9**(1), pp. 62–66, 1979.

[14] J. Kapur, P. Sahoo, and A. Wong, "A new method for gray-level picture thresholding using the entropy of the histogram," *Computer Vision, Graphics, and Image Processing* **29**(3), pp. 273 – 285, 1985.

[15] F. Valverde-Albacete and C. Pelaez-Moreno, "100% classfication accuracy considered harmful:the normalized information transfer factor explains the accuracy paradox," *PLOS ONE* **9**(1), 2014.

[16] F. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms.," in *ICML*, **98**, pp. 445–453, 1998.

[17] T. Fawcett, "An introduction to {ROC} analysis," *Pattern Recognition Letters* **27**(8), pp. 861 – 874, 2006. {ROC} Analysis in Pattern Recognition.

[18] K. A. Spackman, "Signal detection theory: valuable tools for evaluating inductive learning," in *Proceedings of the sixth international workshop on machine learning*, pp. 160–163, Morgan Kaufmann Publishers Inc., 1989.

[19] K. Zuva and T. Zuva, "Evaluation of information retrieval systems," *International journal of computer science & information technology* **4**, pp. 35 – 43, June 2012.

[20] D. A. Jeni, J. F. Cohn, and F. D. L. Torre, "Facing imbalanced data recommendations for the use of performance metrics," in *International Conference on Affective Computing and Intelligent Interaction Workshops*, **15**, pp. 245–251, 2013.

[21] M. Greiner, D. Pfeiffer, and R. Smith, "Principles and practical applications of the Reciever-Operating Characteristic analysis for diagnostic tests," *Preventive Veterinary Medicine* **45**, pp. 23–41, May 2000.

[22] M. Markou and S. Singh, "Novelty detection: A review - part 1: Statistical approaches," *Signal Processing* **83**, p. 2003, 2003.

[23] I. Jolliffe, *Principal Component Analysis*, Springer-Verlag:New York, 1986.

[24] B. Manly, *Multivariate statistical methods, a primer*, Chapman Hall, 2004.

[25] J. J.D., *Applied Multivariate Data Analysis:Categorical and Multivariate Methods*, New York:Springer Verlag, 1992.

[26] P. Rousseeuw, "Multivariate estimation with high breakdown point," *Mathematical Statistics and Applications* **B**, pp. 283–297, 1985.

[27] P. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Society* **79**, pp. 871–880, 1985.

[28] P. Rousseeuw and A. Leroy, *Robust regression and Outlier Detection*, Wiley : New York, 1987.

[29] M. Hubert, P. Rousseeuw, and S. Wan Aeist, "High breakdown robust multivariate methods," *arXiv e-prints* , Aug 2008. Available at https://arxiv.org/abs/0808.0657v1/.

[30] N. Badr and N. Noureldien, "Examining outlier detection performance for principal components analysis method and its robustification methods," *International Journal of Advances in Engineering and Technology* **6**, pp. 573–582, May 2013.

[31] T. Karim, M. Lipu, L. Rahman, and F. Sultana, "Face recognition using pca-based method," in *2010 IEEE International Conference on Advanced Management Science(ICAMS 2010)*, **3**, pp. 158–162, July 2010.

[32] R. Kaur and E. Himanshi, "Face recognition using principal component analysis," in *2015 IEEE International Advance Computing Conference (IACC)*, pp. 585–589, June 2015.

[33] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pp. 586–591, IEEE, 1991.

[34] D. A. Jackson, "Stopping rules in principal components analysis: A comparison of heuristical and statistical approaches," *Ecology* **74**(8), pp. 2204–2214, 1993.

[35] S. Franklin, D. Gibson, P. Robertson, J. Pohlmann, and J. Fralish, "Parallel analysis: a method for determining significant principal components," *Journal of Vegetation Science* **6**(1), pp. 99–106, 1995.

[36] R. Ledesma and P. Valero-Mora, "Determining the number of factors to retain in efa," *Practical Assessment, Research and Evaluation* **12**, 02 2007.

[37] H. Kaiser, "The application of electronic computers to factor analysis," *Ed. Psychol. Meas.* **20**, pp. 141–151, 1960.

[38] R. Gnanadesikan and J. Kettenring, "Robust estimates, residuals and outlier detection with multi-response data," 1972.

[39] T. Bouwmans and E. Zahzah, "Robust PCA via Principal Component Pursuit: A review for a comparative evaluation in video surveillance," *Computer Vision and Image Understanding* **122**, pp. 22–34, May 2014.

[40] D. Donoho, "For most large underdetermined systems of linear equations the minimal l(1)-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics* **59**, pp. 797 – 829, 06 2006.

[41] T. Bouwmans, A. Sobral, S. Javed, S. Jung, and E. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: A review for comparative evaluation with large-scale dataset," *Computer Science Review* **23**, pp. 1–71, February 2017.

[42] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)* **58**(3), p. 11, 2011.

[43] B. S., C. E., and G. M., "Templates for convex cone problems with applications to sparse signal recovery," 2010. Available at http://github.com/cvxr/TFOCS/releases.

[44] D. Laptev, "Robust PCA (Robust Principal Component Analysis) implementation and examples (matlab).," 2014. Available at http://github.com/dlaptev/RobustPCA.

[45] V. C., S. S., P. A.P., and A. S.W., "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization* **21**(2), pp. 572–596, 2011.

[46] R. Paffenroth, K. Kay, and L. Servi, "Robust PCA for anomaly detection in cyber networks," *arXiv e-prints* , Jan 2018. Available at https://arxiv.org/abs/1801.01571/.

[47] C. Holmes, B. Drinkwater, and P. Wilcox, "The post processing of data using the total focusing method," *Insight - Non-Destructive Testing and Condition Monitoring* **46**, pp. 677–680, 11 2004.

[48] E. Villaverde, S. Robert, and P. C., "High frequency total focusing method (tfm) imaging in strongly attenuating materials with the decomposition of the time reversal operator associated with orthogonal coded excitations," in *AIP Conference Proceedings 1806*, American Institute of Physics, 2017.