Weld Defect Detection using Ultrasonic Phased Arrays

by

Bryan Cassels

A thesis submitted in partial fulfilment for the requirements for the degree of Doctor of Philosophy at the University of Central Lancashire in collaboration with BAE Systems Maritime

September 2018

STUDENT DECLARATION FORM

Type of Award   **Doctor of Philosophy**

School   **Engineering**

1.   Concurrent registration for two or more academic awards

I declare that while registered as a candidate for the research degree, I have not been a registered candidate or enrolled student for another award of the University or other academic or professional institution

2.   Material submitted for another award

I declare that no material contained in the thesis has been used in any other submission for an academic award and is solely my own work
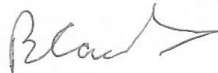
3.   Collaboration

Where a candidate's research programme is part of a collaborative project, the thesis must indicate in addition clearly the candidate's individual contribution and the extent of the collaboration. Please state below:

The project is in collaboration with the sponsor BAE Systems Maritime. They have provided funding, access to experimental facilities and advice. However, all the work in this thesis is from my own investigations and evaluations.

4.   Use of a Proof-reader

No proof-reading service was used in the compilation of this thesis.

Signature of Candidate

Print name: Bryan Cassels

Abstract

Many traditional ultrasonic test methods, based on the manipulation of an ultrasonic probe by an experienced inspector, are beginning to be replaced by Automated Ultrasonic Testing (AUT). Largely limited to regular structures, the integration of phased arrays and computer controlled mechanical manipulators allows AUT to provide fast, regular and repeatable data acquisitions for off-line inspection and future auditing. The objective of this thesis is to investigate methods of assisting with the inspection of these vast quantities of data. To this end the emphasis is on detecting regions of a weld that do not comply to the normal, anomaly free, structure.

It is found that a multivariate analysis, using Principal Component Analysis (PCA), improves the Probability Of Detection (POD) of anomalies over univariate techniques, which rely on only the segmentation of regions with unusually high ultrasonic reflections. A further finding is that the multivariate approach is capable of mitigating the effects of dominating front wall reflections and other continuous features.

Experimental results using test block data reveal a high POD with a low false alarm rate. This is particularly the case where the probe is in direct contact with the test piece and the front wall is gated out. Despite a lower POD the immersion results are particularly significant in that they permit the use of segmentation procedures simply not possible in the univariate case.

Although limited to 2D anomaly location it is relatively straightforward, using the 2D location as a key, to extract full volumetric data of the anomaly from the original 3D ultrasonic data set. A future extension of this work is to use the 3D information to accurately classify any anomaly. In addition to providing a more detailed description of the anomaly this also has the potential to reduce the false alarm rate.

AUT produces vast quantities of data for inspection. It is increasingly common for this data to be in the form of a sequence of images representative of a weld's cross section. Manual inspection of each image, by qualified personnel, is both expensive and prone to human error. The system developed here has the potential to improve this process by considerably reducing inspection time, and cost, whilst maintaining a consistently high POD.

Acknowledgements

I would first like to give my thanks to my supervisors Lik-Kwan Shark and Stephen Mein without whose help much of the work presented in the thesis could not be done. They both provided inspirational ideas with valuable insights and continual enthusiasm.

Sincere thanks are also expressed to Tom Barber, Andrew Nixon and Ray Turner for their valuable advice. In addition to giving considerable encouragement throughout the entire programme they also helped with experiments and provided data without which, this work would largely contain only simulations.

For project funding acknowledgement and thanks must be given to BAE Systems Maritime.

Finally I must give thanks to my wife, Pat, for her endless patience, encouragement and support. Not forgetting Dylan and Rosie who remind me of the important sides to life.

Bryan Cassels

August 2018

Contents

Abbreviations

ADMM	Alternating Direction Method of Multipliers
AUC	Area Under Curve
AUT	Automated Ultrasonic Testing
BIP	Beam Intersection Point
FAR	False Alarm Rate
FN	False Negative
FP	False Positive
FBH	Flat Bottomed Hole
FMC	Full Matrix Capture
GT	Ground Truth
HDLSS	High Dimensional Low Sample Size
KI	Kittler and Illingworth (Minimum error thresholding)
LOF	Local Outlier Factor
MCD	Maximum Covariant Determinant
MCE	Miss-Classification Error
MCC	Mathews Correlation Coefficient
ME	Maximum Entropy (thresholding)
MVE	Maximum Volume Ellipsoid
NF	Near Field

NPD	Negative Predictive Value
PA	Phased Array
PC	Principal Component
PCA	Principal Component Analysis
PCP	Principal Component Pursuit
POD	Probability Of Detection
ROC	Receiver Operating Characteristic
RPCA	Robust Principal Component Analysis
SAFT	Synthetic Aperture Focusing Technique
SBH	Spherical Bottomed Hole
SDH	Side Drilled Hole
SNR	Signal to Noise Ratio
TB	Test Block
TFOCS	Templates for First Order Conic Solvers
TN	True Negative
TP	True Positive
TFM	Total Focusing Method

Chapter 1

Introduction

Welding is perhaps the most widely used method of joining together metals and alloys. Since the beginning of the 20th century with the formation of organisations such as the American Welding Society, documented research into welding has continued to accelerate. A significant result is the development of numerous processes and methods of automation. Today many techniques exist that are capable of efficiently and continuously producing welds with a high degree of reliability. Similarly quality assurance techniques are mature and encompass control of the whole process. For example BS EN ISO 3834 [1] sets out requirements of weld quality for manufacturers to meet; BS 3923-1 [2] sets out methods of inspection, whilst standards such as Defence Standard 02-773 [3] apply to specific environments. Conformance to these standards comes at considerable expense through the use of highly qualified personnel and time. Justification is that the effect of just a single weld failure can be catastrophic. For example when many large structures (ranging through buildings, bridges, dams, pipelines, ships' hulls, aircraft, nuclear power plants and power generator boilers) depend on the integrity of their welds, a single failure has potential for high monetary costs, damage to the environment or loss of human life [4], [5]. Consequently the expectation of weld quality has never been higher.

Following visual examination of a weld, one of the earliest forms of Non Destructive Testing (NDT) was the use of a liquid penetrant dye [6]. The earliest test, known as the 'Oil and Whitting Method', was introduced in the late 19th century. However it was not until shortly after the sinking of the Titanic in 1912, that the roots of modern NDT and Non Destructive Evaluation (NDE) began to be established [5]. In the 1940's the accepted standard of inspection for critical structural welds was, according to DeNale and Lebowitz [7], radiography. They also record that the gradual introduction of ultrasonic inspection,

1

for similar applications, only started in the early 1960's. Over much of the intervening time radiography remained, for safety critical applications, the preferred method.

In principle, the radiographic test procedure is straightforward. On one side of the weld is a film (or more recently an electronic detector) and on the other is the source of radiation. As the beam of radiation passes through the weld, its interaction with anomalies causes higher levels of attenuation than does that through the more homogeneous, anomaly free, background. After developing, the film contains an image of the weld, and its surrounding area (figure 1.1). This provides a permanent record of the weld's internal structure. It can (depending on film's size) cover a long length of weld. To distinguish this type of projection onto film, from that of a photograph produced by light, the image is known as a radiograph.

There are, however, a number of difficulties with radiography, the most serious being due to radioactivity. Not only must strict health and safety rules be followed during operation but an entire work area may need to be evacuated before tests begin and for a period afterwards. This is costly in terms of lost production. In addition to this expense other limitations of radiography are that it is highly directional, sensitive to the orientation of a flaw, does not indicate the depth of the flaw and requires a high degree of skill and experience for exposure and interpretation [7].
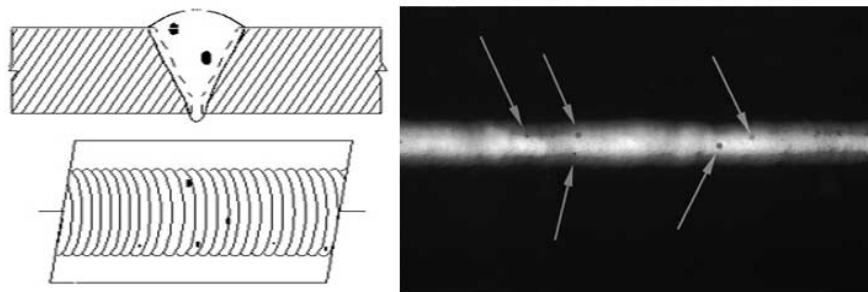
Figure 1.1: Example of a radiographic recording (from www.nde-ed.org)

Ultrasonic inspection using a single element probe does not suffer from these health and safety limitations. Through a sequence of adaptive closed loop manual operations an inspector, through skill and experience, is able to detect and reliably sentence any anomaly within the weld. However such approaches remain 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.

1.1 Industrial application

In the early 1960's researchers started to develop ultrasonic phased arrays. In contrast to the single element probe these initially contained a line of small equally spaced point sources. As each element requires a dedicated channel to excite and record the response from each ultrasonic element, complete systems were large and expensive. Continuing developments in micro-electronic lithography meant that, starting around the 1990s, low power CMOS technology had developed to the extent that inexpensive and portable multi-channel phased array systems were gaining in popularity. A significant advantage of these systems over single element probes is their ability to apply a timed sequence of pulses to individual elements. This allows the ultrasonic beam to be steered and focused, as in the case of a sectorial scan. Alternatively it is possible to create an image through a process of synthetic focusing at each pixel point in an image. Examples of this are the Synthetic Aperture Focusing Technique (SAFT), [8] and the Total Focusing Method (TFM), [9].

a) Sector Scan   b) TFM Image
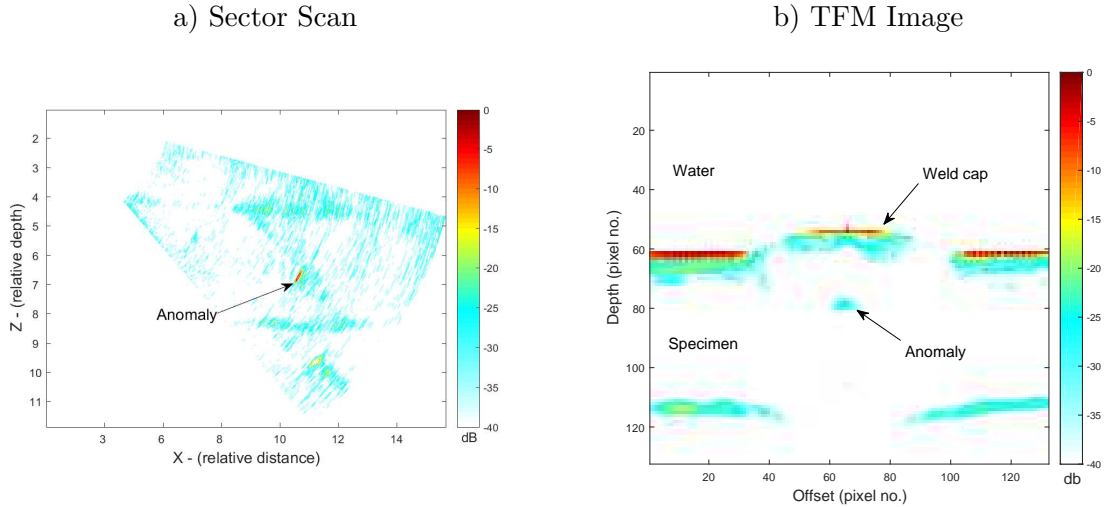
Figure 1.2: Example cross sectional images using a phased array

The two imaging methods of interest to this work (both of which are described in the early chapters) are, in fact, those due to sectorial scanning and TFM. Examples of each are illustrated in figure 1.2. For purposes of illustration the instrument gain for the sector scan is set high. One result of this is that the fusion face of the 'double V' butt weld is evident by a reversed 'Z' pattern ($\mathrm{Ƨ}$). In the case of this particular sectorial image the probe is in contact with the test piece, and to its side. This is unlike the TFM example where the immersion probe is mounted directly above the test piece; acoustic coupling between the

probe and test piece is via the intervening water layer. At present no further explanation of these set ups or phased array operation is given. It suffices only to indicate that such images are possible.

Cross-sectional images provide information on the depth and size of an anomaly at the particular index point along the weld. However this only represents a thin 'slice' of the material within the phased array's 2-dimensional active zone. To cover a length of weld, as in the case of a radiograph, a number of adjacent ultrasonic images are required. The smaller the distance between each cross-section the finer is the resolution. To determine the size of anomalies in this the step size from image to image must be carefully controlled. The smaller the step size the finer is the resolution. Maintaining a constant path along the weld is usually best achieved by mounting the phased array into a manipulator designed for the specific application (examples given in figure 1.3 are illustrative only). Depending on the amount of time required for each data acquisition an encoder may be used to trigger an acquisition at each index point as the carrier is moved, manually or by electric motor, along the weld. Alternatively, for longer data acquisition times, the carrier is stepped to the next index point after each acquisition.

(Olympus IMS)   (GE Krautkramer Weldstar)

Figure 1.3: Example probe carrier systems for AUT

Driven by a worldwide demand for the distribution of gas, oil and water there is now a long history of high productivity pipeline welding [10]. A common feature of the resulting girth welds is that they are ideally suited to Automated Ultrasonic Testing (AUT). In this context, and using probes mounted on carriers similar to those depicted in figure 1.3, AUT itself now has a similarly long history [11].

AUT data of this type continues to be acquired through the use of single element probes using pulse-echo or pitch-catch techniques [12] with inspection covering specific zones of

the weld [13]. Using phased arrays data capture need not be limited to zones. Instead, by recording full cross sectional images at regular index positions volumetric information covering the entire weld is recorded, thereby potentially improving the probability of detecting anomalies anywhere within the scanned volume.

1.2   Potential benefit

A downside to AUT is that each image now needs to be inspected individually. Whilst an inspector's experience can compensate for the negative effect of time pressure and mental workload, [14], other studies of human factors affecting the reliability of manual inspection [15], [16], 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 will only add to the cost and demands in terms of time and mental workload placed on inspectors.

To assist with the inspection of these 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 [17], [18]. Today this theme continues [19], [20] and the indications are that research remains largely directed towards the classification and sizing of individual defects. Although of value in classifying anomalies once found, 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.

1.3   Outline of thesis

As alluded to above, numerous papers investigating methods of anomaly identification and sizing have been published over the last 30 years. In the background review for this

proposal no literature attempting to address the problem of anomaly location within large sections of a weld was found. Consequently this becomes the broad objective of this work.

The main assumptions of this work is that the weld is linear. Examples of this include the circumference of a pipe or a long butt-weld between two plates. There are no sharp angles or other geometric irregularities. As this is thought to be the first work of this kind this is a reasonable starting constraint. The significance is that the anomaly free background is constant and that data acquisition can be automated. Problems of less regular structures become a subject of future work.

Automatic data capture provides a number of identical acquisitions at a set of identically spaced index points. A statistical description of the anomaly free condition provides a reference to which new acquisitions can be compared. This does not, however, imply that the exercise is to be treated as a purely statistical or black box problem. Although for this work the transducers are the same type (linear phased array) there are different methods of operation, different methods of data capture and different methods of image creation. In short any study of this nature requires at least some appreciation of the fundamentals of ultrasonic transmission and reception. Briefly the organisation of this thesis follows a set of logical steps starting with the fundamentals of ultrasonic transmission through a basic description of the ultrasonic phased array to methods of providing data for inspection. Statistical methods of analysing the data are followed by the presentation of results and an evaluation of the techniques considered. A more specific organisation of this thesis is given by the following list:-

1. **Ultrasonic principles, transducers, data acquisition and presentation**

   Chapter 2 provides an overview of the ultrasonic wave equation and introduces some of the principles of wave propagation required for an understanding of the operation and interpretation of data from ultrasonic transducers. Next, chapter 3 discusses the operation of ultrasonic transducers in more detail with particular emphasis on the linear phased array. This includes a discussion of beam steering and focusing as well as the focal law calculations, data capture and imaging.

2. **Data sets**

   Details of the test blocks from which experimental data is acquired are given in chapter 4. In addition to the dimensions of each test piece information regarding the location, size and type of anomaly is given. In some cases this data is not exact and test pieces are found to contain unintentional anomalies. The chapter, therefore, outlines the methods of establishing ground truth information for each test specimen.

6

To automate the tests the ground truth data is included in a test vector which is, in turn, used by a test bench to compare true regions of anomalies with those detected, thereby enabling the creation of confusion statistics for evaluation.

3. **Thresholding**

The simplest method of anomaly detection is that of thresholding. Chapter 5 discusses methods of first creating a suitable image from captured data. It then investigates various methods of establishing a threshold and presents results. The methods adopted here are, however, only suitable where front wall echoes are not present. This is not always possible, for example in the case of immersion tests or where near surface defects are to be detected.

4. **Principal Component Analysis (PCA)**

Ultrasonic data capture combining sets of data acquisitions from different elements of a multi element probe is, by its nature, multidimensional. Principal Component Analysis (PCA) is perhaps the simplest way of analysing multidimensional data. Although PCA has been applied to ultrasonic data in the past this has tended to be for specific purposes such as anomaly classification [21]. This is thought to be the first time that PCA has been applied to the detection of anomalies in large quantities of weld data. The introduction to PCA (chapter 6) demonstrates, and provides results, of the technique using manually selected observations; that is, observations considered to be free from anomaly and hence representative of background.

5. **Robust PCA (RPCA)**

A recognised problem of PCA is its susceptibility to outliers in the data set. Manual selection to avoid outliers, as done previously, would exclude PCA as a technique for automatic anomaly detection. To overcome the problem of outliers two approaches are compared. These are by trimming and Principal Component Pursuit (PCP). Although both fall under the general heading of RPCA this term is, here, used only to describe the method due to PCP. The first approach is referred to by its descriptive term of 'trimming'. Chapter 7 describes trimming whilst RPCA by PCP is the subject of chapter 8.

6. **Conclusions**

The final chapter provides an overview of results presented in chapters 6 to 8. This is followed by a review of the thesis in terms of its contribution to knowledge, recommendations for future work and comments on its industrial impact.

Chapter 2

Properties of ultrasonic signals

As outlined in the introduction the emphasis of this work is to detect anomalies in welds using ultrasonic techniques. Although these primary objectives are to be addressed through the analysis of resulting images, rather than the ultrasound itself, it is important to understand how images are created. This requires some knowledge both of the data acquisition and, in turn, how the transducers operate.

Even with knowledge of transducer operation, in practice it remains possible for images to contain features, or artifacts, that result from the way the ultrasound itself propagates through the actual material. In this context this chapter presents an overview of ultrasound. It introduces the mechanisms by which ultrasound propagates through a material, how it interacts at boundary interfaces, how mode conversions occur and how its intensity reduces with distance. A description of the transducers, their characteristics and mode of operation is left to the next chapter.

2.1 Ultrasound

In a general sense ultrasound is a mechanical wave with a frequency above the upper audible limit of human hearing (20 kHz). Over an extended range of intensities these waves have a wide variety of applications. High intensity applications, for example, include the cutting and cleaning of material; at lower intensities ultrasonic waves include sonar, medical and non-destructive testing. This chapter discusses the fundamental principles of ultrasound as it pertains to non-destructive testing. Here, ultrasonic waves provide a mechanism for both detecting the presence of anomalies within a solid material (detection) and providing

an indication of their characteristics (characterisation).

For this work ultrasonic testing involves the use of electro-acoustic transducers which act as both a transmitter and receiver of sound. During transmission, an electrical impulse causes the transducer's face to vibrate. When receiving, any displacement of the transducer's face is converted to a corresponding electrical signal. Most transducers are capable of both conversions with equal efficiency. The following discussion introduces the fundamental physics of sound transmission in solids. Initially the assumption is that sound energy propagates as a series of parallel compressions and rarefactions. In particular the discussion assumes plane wave motion; that is, the transducer face moves in a piston-like manner, the phase of the wave across any plane parallel to the transducer's surface being constant. Higher frequencies permit the detection of defects with smaller dimensions. However, as the wavelength reduces to that of the dimensions of the material's grain structure, the wave's attenuation with distance becomes significant. In the case of ultrasonic testing of metallic materials frequencies tend to be in the range of 2 MHz to 30 MHz, the particular frequency being a compromise between depth of test region and resolution.

2.2 The travelling wave

The most general model of ultrasonic propagation within an elastic medium is provided by the 3 dimensional wave equation. However, for clarity, a one dimensional model provides a much simpler basis on which to introduce the basic concepts and terminology used throughout this thesis.

If an ultrasonic transducer is in direct contact with a test piece and vibrating with simple harmonic motion and constant amplitude then the particles at the surface of the test piece will vibrate in an identical pattern. Assuming a peak amplitude of y_m and a frequency $f = \omega/2\pi$ the surface particles, those at depth $x = 0$, will displace from their mean position, y, according to:-

$$y = y_m sin\omega t$$

The wave will now propagate into the depth of the material with constant velocity v, the time (t') for the wave to travel a distance x from the source being equal to x/v. Consequently the phase of the wave at point x lags that at the surface, where $x = 0$, by an amount $\omega t'$ and the vibration at that point is expressed as:-

$$y = y_m sin\omega(t - t') \quad or \quad y = y_m sin\omega(t - x/v)$$

9

This gives the same waveform at the point $x = vt$ at time t, as was present at $x = 0$ at time $t = 0$. To follow a phase of the wave as time progresses, $t - x/v$ has a fixed value. Consequently as t increases, x must also increase, illustrating that v is the phase velocity of the wave.

A single time period T, during which the phase angle changes by 2π, defines the wavelength (λ), where:-

$$\lambda = vT$$

and

$$y = y_m \sin 2\pi((t/T) - (x/\lambda)) \qquad = \qquad y_m \sin(\omega t - kx) \qquad (2.1)$$

where k, the wave number, is defined as:-

$$k = 2\pi/\lambda \quad = \quad \omega/v \qquad (2.2)$$

The equation of the harmonic travelling wave is sometimes written as:-

$$y_m \sin(kx - \omega t) \qquad (2.3)$$

The two are actually the same with the exception of a $180°$ phase difference ($sin(\omega t - kx) = sin(-(kx - \omega t)) = -sin(kx - \omega t)$).

2.3 Particle velocity

Associated with particle displacement, y, is a particle velocity, u, and a change in acoustic pressure, p. For particle velocity:-

$$u = \frac{\partial y}{\partial t} = u_0 cos(\omega t - kx) \qquad (2.4)$$

where $y_m \omega$ has been replaced by the particle velocity amplitude, u_0. This also demonstrates that the particle velocity amplitude leads the particle displacement by $90°$.

Like displacement, particle velocity also has a harmonic form. Before considering acoustic pressure, p, which also has a harmonic form, it is worth introducing the general wave equation.

2.4 The general wave equation

With appropriate zeroing of the scales for x and t it is possible for the phase functions to be written as either [22]:-

$$sin(\omega t - kx), \quad cos(\omega t - kx) \quad or \quad e^{j(\omega t - kx)}$$

The use of exponentials has the advantage of simplifying differentiation and integration. Letting Φ represent either y, u, or p it is possible to write:-

$$\Phi = \Phi_0 e^{j(\omega t - kx)} \qquad (2.5)$$

Differentiating 2.5 twice with respect to t and k results in:-

$$\frac{\partial^2 \Phi}{\partial x^2} = -k^2 \Phi_0 e^{j(\omega t - kx)},$$

and

$$\frac{\partial^2 \Phi}{\partial t^2} = -\omega^2 \Phi_0 e^{j(\omega t - kx)},$$

leading to:-

$$\frac{\partial^2 \Phi}{\partial t^2} = \frac{\omega^2}{k^2} \frac{\partial^2 \Phi}{\partial x^2} \qquad (2.6)$$

from equation 2.2:-

$$\frac{\partial^2 \Phi}{\partial t^2} = v^2 \frac{\partial^2 \Phi}{\partial x^2} \qquad (2.7)$$

This is called the general equation for plane waves [22]. As suggested Φ represents any wave characteristic such as y, u and p. For an elastic medium acoustic pressure is the result of the motion of particles about their equilibrium position. The resulting compressions and rarefactions cause changes in material density as energy is transferred in the direction of the travelling wave. It is therefore apparent that p is dependent on the characteristics of the medium. To investigate the material parameters affecting p an alternative approach of deriving the general wave equation, based on a physical model, is required.

Taking into account the x, y, and z directions equation 2.7 can be extended to:-

$$\frac{\partial^2 \Phi}{\partial t^2} = v^2 \left(\frac{\partial^2 \Phi}{\partial x^2} \quad + \quad \frac{\partial^2 \Phi}{\partial y^2} \quad + \quad \frac{\partial^2 \Phi}{\partial z^2} \right)$$

or, using the Laplace operator:-

$$\frac{\partial^2 \Phi}{\partial t^2} = v^2 \nabla^2 \Phi \tag{2.8}$$

2.5 Alternative derivation of the general wave equation

Acoustic pressure is introduced by first considering a physical model of the one dimensional wave equation. Within the material's elastic limit the force to displace a particle can be determined in two ways. Firstly Newton's second law [23] defines the force to be the product of the particle mass (m) and its acceleration (a):-

$$F_{Newton} = ma$$

Secondly, Hooke's law [23]:-

$$F_{Hooke} = -k'x$$

defines the displacement x to be proportional to the external force. The constant of proportionality, k', is a characteristic of the material's stiffness. Using the conservation of energy the wave equation is now derived by equating Newton's second law to Hooke's law. Before further consideration it is useful to make an analogy between particle movement and a mass-spring system, figure 2.1.

Here $\Phi(x)$ is the displacement of the mass m from its equilibrium position x. The forces exerted on the centre m, at position $(x + \Delta x)$ are, according to Newton:-

$$F_{Newton} = ma = m\frac{\partial^2}{\partial t^2}\Phi(x + \Delta x, t)$$

and according to Hooke:-

$$F_{Hooke} = k'[\Phi(x + 2\Delta x, t) - \Phi(x + \Delta x, t)] + k'[\Phi(x, t) - \Phi(x + \Delta x, t)]$$
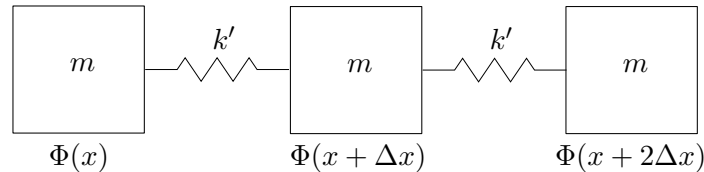
Figure 2.1: Particle displacement as a mass spring system

Equating these two forces gives

$$m\frac{\partial^2}{\partial t^2}\Phi(x+\Delta x,t) = k'[\Phi(x+2\Delta x,t) - \Phi(x+\Delta x,t)] + k'[\Phi(x,t) - \Phi(x+\Delta x,t)]$$

With N particles equally spaced over a length L such that $L = N\Delta x$, the total mass is $M = Nm$ and the total spring stiffness is $K = k'/N$. The previous equation can now be written as:-

$$\frac{\partial^2}{\partial t^2}\Phi(x+\Delta x,t) = \frac{KL^2}{M}\frac{[\Phi(x+2\Delta x,t) - 2(\Phi(x+\Delta x,t) + \Phi(x,t)]}{\Delta x^2}$$

In the limit as $\Delta x \to 0$

$$\frac{\partial^2 \Phi}{\partial t^2} = \frac{KL^2}{M}\frac{\partial^2 \Phi}{\partial x^2} \tag{2.9}$$

This is equivalent to equation 2.7 where the term $\frac{KL^2}{M}$ represents v^2. This can be confirmed by fundamental dimensional analysis where the term $\frac{KL^2}{M}$ has dimensions $\frac{ML^2}{T^2M}$ which is the same as v^2.

2.6 Acoustic pressure

Continuing with the mass spring system analogy any spatial disturbance results in compression or expansion of individual masses leading to an explanation of the dependence of pressure on displacement. Consider the two opposite sides of an elemental mass, at rest, to be x_1 and x_2. If an elemental distance $(x_1 - x_2)$ is made much less than the wavelength of the sound wave then the undisturbed volume, (V), of the mass element with cross sectional area A is:-

$$V = A(x_2 - x_1)$$

When subject to a disturbance x_1 moves to $x_1 + y_1$ and x_2 moves to $x_2 + y_2$. The new volume becomes:-

$$V + \delta V = A(x_2 + y_2 - x_1 - y_1)$$

so that:-

$$\delta V = A(y_2 - y_1)$$

Under compression the density of the mass will increase, whilst under rarefaction it will reduce. The ratio between the pressure change, p, and the proportional volume change is the material's elastic constant, K. The pressure change in response to a volumetric change

13

is, therefore:-

$$p = -K\frac{\delta V}{V} \quad = \quad -K\frac{(y_2 - y_1)}{(x_2 - x_1)}$$

In the limit where $x_2 - x_1$ is very small this gives the relationship between displacement y and pressure, p as:-

$$p = -K\frac{\partial y}{\partial x} \tag{2.10}$$

This demonstrates that acoustic pressure, like particle velocity u (equation 2.4), leads particle displacement by 90°.

2.7 Wave modes

For a solid the velocity of the wave is in general (Equation 2.9 and [12]), dependent on the material's density, ρ, and elastic constant, K.

$$v = \sqrt{\frac{K}{\rho}} \tag{2.11}$$

Here K represents a volumetric elastic constant. For a longitudinal wave, where the absolute direction of the particle is in the same direction as the wave, then assuming a constant cross sectional area, the elastic constant refers to Young's modulus, E. Hence the velocity of a longitudinal wave is written as:-

$$v_L = \sqrt{\frac{E}{\rho}} \tag{2.12}$$

and for the velocity of a shear wave:-

$$v_S = \sqrt{\frac{G}{\rho}} \tag{2.13}$$

where G, the modulus of rigidity, is defined as ratio of shear stress to shear strain.

Ultrasonic testing is usually carried out with frequencies in the Mega-Hertz so that the wavelength is much smaller than the dimensions of the object under test. In a compressible material such as steel the longitudinal stresses lead to a lateral longitudinal compression or stretching of the material. According to the law of conservation of mass the material now exhibits a change in cross sectional area (due to strain). This phenomenon is called

14

the Poisson effect. Poisson's ratio σ, is a measure of the fraction (or percent) of expansion divided by the fraction (or percent) of compression, for small values of these changes.

Young's modulus and the modulus of rigidity are related by:-

$$G = \frac{E}{2(1 + \sigma)} \tag{2.14}$$

Taking into account the Poisson effect the velocity of a longitudinal wave now becomes, [12]:-

$$v_L = \sqrt{\frac{E(1 - \sigma)}{\rho(1 + \sigma)(1 - 2\sigma)}} \tag{2.15}$$

and for a shear wave:-

$$v_S = \sqrt{\frac{E}{2\rho(1 + \sigma)}} \tag{2.16}$$

For many practical purposes the general equation (2.12) is adequate for determining longitudinal velocity with equation (2.13) being adequate for shear velocity, [12].

2.8   Specific acoustic impedance

Using equation 2.4 it is possible to provide a relationship between particle displacement, y, and particle velocity, u. This is:-

$$u = \frac{\partial y}{\partial t} = -y_m \omega cos(kx - \omega t)$$

Similarly pressure 2.10 may be represented as:-

$$p = -K\frac{\partial y}{\partial x} = -Ky_m k cos(kx - \omega t)$$

Specific acoustic impedance, z, is defined by p/u and, from the above, may be written as:-

$$z = Kk/\omega$$

After substitution and re-arrangement from equations 2.11 ($K = \rho v^2$) and 2.2 ($\omega/k = v$) specific acoustic impedance is written as:-

$$z = \rho v \tag{2.17}$$

This is a highly useful concept in ultrasonics and there is a direct analogy with electrical circuits where for maximum power transfer between two circuits the impedances must match so that there is minimal reflection. Ultrasonic testing exploits the fact that a reflection, caused by a change in acoustic impedance, is the result of a fault or discontinuity.

2.9 Reflection and transmission of plane waves

For a homogeneous medium of constant density the velocity and direction of a plane wave remains constant. Any change in density results in a change in acoustic impedance so that part of the pressure wave is reflected. Figure 2.2 depicts a common ultrasonic inspection situation where a test piece ($M2$) is coupled to an ultrasonic transducer by a coupling agent such as water ($M1$). The displacement of an incident wave varies according to:-

$$y_i = y_m \sin(k_1 x - \omega t) \tag{2.18}$$

At normal incidence to the interface (figure 2.2a) the direction of the reflection is $180°$ so that the reflected displacement varies according to:-

$$y_r = -y_m \sin[-(k_1 x + \omega t)] \tag{2.19}$$

whilst the remaining energy transmits with no change of direction:-

$$y_t = y_m \sin(k_2 x - \omega t) \tag{2.20}$$

At oblique incidence (Figure 2.2b) the angle of the reflection is symmetrical with the angle of incidence about the normal (*ie.* $\theta_i = \theta_r$) whilst the transmission refracts according to Snell's law (appendix B):-

$$\frac{\sin(\theta_i)}{v_1} = \frac{\sin(\theta_r)}{v_2} \tag{2.21}$$

At normal incidence both the reflected and transmitted waves are longitudinal. The proportion of each component depends on the respective acoustic impedance (Z_1 and Z_2) of the two materials (M_1 and M_2) forming the interface. For normal incidence the proportions of the pressure amplitude reflected (the reflection coefficient, R) and that transmitted (the transmission coefficient, T) are [12]:-

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{2.22}$$

16

(a) Normal Incidence   (b) Oblique Incidence
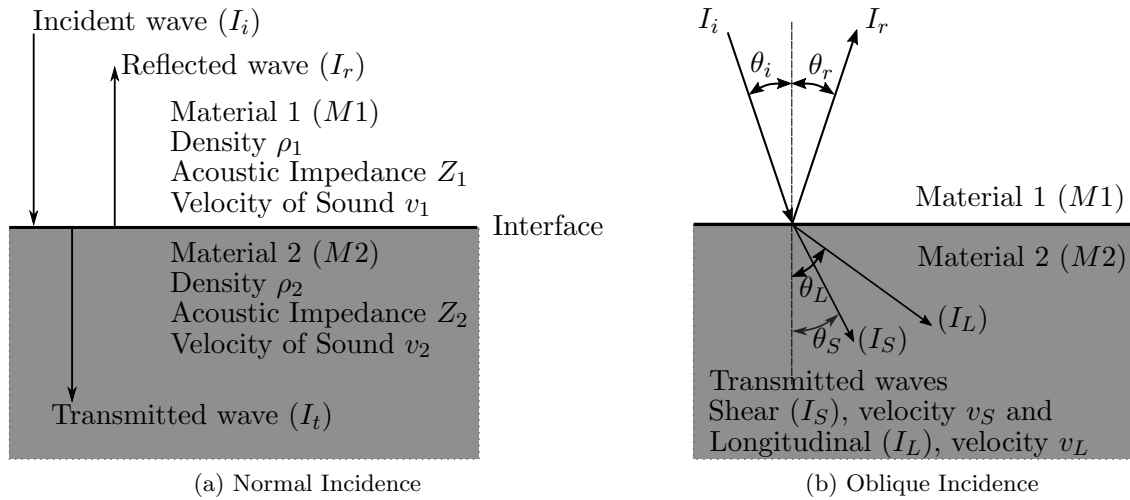
Figure 2.2: Waves at normal and oblique incidence

$$T = \frac{2Z_2}{Z_2 + Z_1}$$ (2.23)

For most ultrasonic inspections $M2$ will be a solid, such as steel, supporting both longitudinal and shear wave propagation. At normal incidence $M2$ is subject only to a tensile force so that the transmitted wave is dominantly longitudinal. As the angle of incidence deviates from the normal $M2$ becomes subject to an additional shear force. Different longitudinal and shear wave velocities (equations 2.12 and 2.13) lead to two different angles of refraction. Z_1 must now account for the angle of incidence and in addition Z_2 must account for both the shear and the longitudinal waves. Krautkramer and Krautkramer [12] state the following equations for each impedance:-

$$Z_1 = \frac{\rho_1 v_1}{cos(\theta_i)}$$

$$Z_2 = Z_L cos^2(2\theta_S) + Z_S sin^2(2\theta_S)$$

where

$$Z_L = \frac{\rho_2 v_L}{cos(\theta_L)}$$

and

$$Z_S = \frac{\rho_2 v_S}{cos(\theta_S)}$$

17

With these modifications the equations for reflection and transmission coefficients are [12]:-

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{2.24}$$

$$T_L = \frac{\rho_1}{\rho_2} \frac{2Z_L cos(2\theta_S)}{Z_2 + Z_1} \tag{2.25}$$

$$T_S = \frac{\rho_1}{\rho_2} \frac{2Z_S sin(2\theta_S)}{Z_2 + Z_1} \tag{2.26}$$

2.10   Echo transmittance and critical angles

For an inspection it is desirable to transmit as much of the ultrasonic wave into the test material as possible and then to receive the maximum possible echo. For a water-steel interface equation 2.23 reveals a transmission coefficient of 1.938 (assuming a velocity and density for steel of $5890m/s$. and $7870Kg/m^3$ respectively). Any echo takes the return path through the steel-water interface where the transmission coefficient is now 0.0618. The product of both transmission coefficients gives the echo transmittance, in this case 0.1226. This figure represents the maximum amount of pressure energy returned from the original sound wave at normal incidence.

For any other angle of incidence the echo transmittance is more difficult to determine. In particular, for the water-steel interface, the transmitted wave now has two components so that equation 2.23 is replaced by equations 2.26 and 2.25. A further complication is the fact that under certain angular ranges the equations do not have real solutions, for example when the transmitted wave has an angle of $90°$ and above. In general, therefore, the solutions to the equations are complex. Figures 2.3.a and b are simplified by not including angular solutions with complex values (for example a transmitted wave with $90°$, or more, of refraction).

When the incident wave is normal to $M2$ (figure 2.2a) the transmission coefficient of the shear wave is zero (equation 2.26) whilst that of the longitudinal wave is maximum (equation 2.25). As the angle of incidence increases, the vectors representing the transmitted longitudinal and shear waves change in direction and magnitude. Both components refract, with the longitudinal wave leading that of the shear wave. The difference is explained by Snell's law and the differences in v_L and v_S. The differences in magnitude are explained by the equations for the transmission coefficients (equations 2.26 and 2.25). At some angle of

18

a) Water-steel ($steel - \rho = 7870 Kg/m^3, v_L = 5890 m/s, v_S = 3250 m/s$)
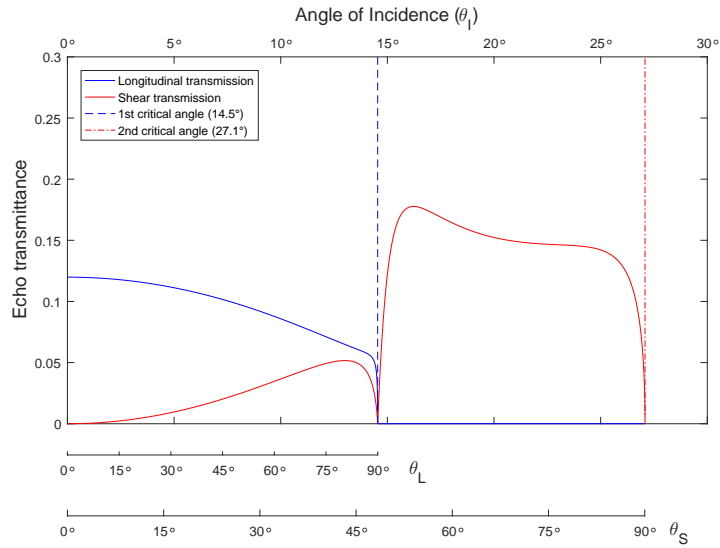
b) Rexolite-steel ($Rexolite - \rho = 1050 Kg/m^3, v_L = 2350 m/s, v_S = 1555 m/s$)
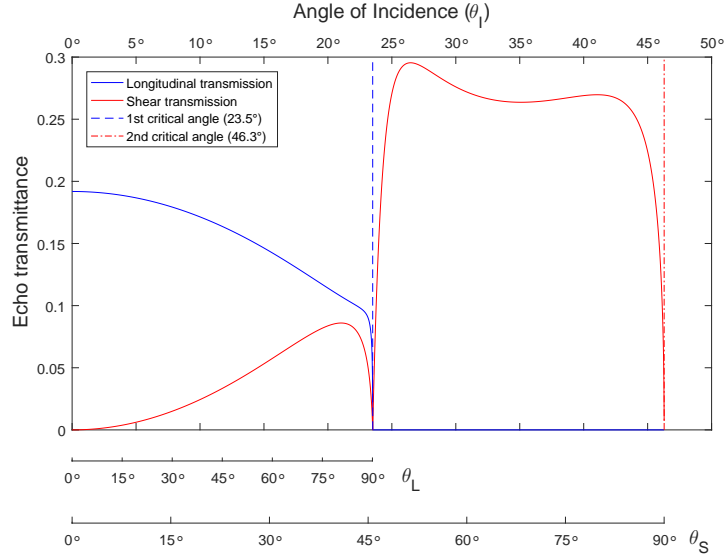
Figure 2.3: Echo transmittances for longitudinal incident waves

19

incidence the longitudinal wave is refracted by 90° and its magnitude reduces to zero. This is the first critical angle. As the angle of incidence increases further the energy balance is maintained by a surface wave and the refracted shear wave. The angle at which the shear wave refracts by 90° is known as the second critical angle.

Beyond each critical angle the solutions contain imaginary values. The figures show only real values. The imaginary components dissipate as surface waves [12]

2.11   Acoustic energy density

Acoustic energy density (E) is a measure of the amount of sound energy contained in a unit volume of a planar wave. It is the sum of kinetic (E_k) and potential (E_p) energy. A particle or small volume V_0 of material with a mass of $\rho_0 V_0$ and moving with velocity v has kinetic energy:-

$$E_k = \frac{1}{2}\rho_0 V_0 v^2 \tag{2.27}$$

Any change in volume from V_0 to V due to compression representing a change in potential energy is:-

$$E_p = -\int_{V_0}^{V} p dV \tag{2.28}$$

The negative sign indicates that the potential energy increases when the volume decreases due to a positive acoustic pressure p.

2.12   Acoustic intensity

Acoustic intensity (I) is a measure of the energy per unit area in a planar sound field; it remains constant at all points in the sound field irrespective of distance. The units are W/m^2 and may be determined from the pressure and acoustic impedance ([12] or by analogy with electrical circuit theory) as:-

$$I = \frac{p^2}{Z} \tag{2.29}$$

At a boundary there is an energy balance between the incident acoustic intensity and the resultants. In the case of the water to steel interface (e.g. $M1$ to $M2$ figure 2.2) the fractional component of each resultant intensity (I_L, I_S, I_r), is in accordance with equations 2.24 to 2.26.

20

In section 2.10 the transmission coefficient between water and steel was calculated as 1.938, or 193.8%. At face value this appears to contradict the law of conservation of energy and the above energy balance. However the acoustic impedance of steel is more than 30 times that of water so despite the increase in pressure the intensity of the transmitted wave in $M2$ (steel) is much smaller than that in $M1$ (water).

2.13   Attenuation

As a sound wave propagates through a medium energy is, in fact, lost with distance. This attenuation is due to a combination of scattering and absorption [12]. In most metals scattering results from the fact that the material is polycrystalline in nature with grain boundaries creating sudden changes in acoustic impedance. The amount of scattering is largely dependent on the ratio of the sound wave's wavelength to that of the average grain size of the material. If the wavelength is similar to or smaller than the grain size, then at each boundary it may split into a variety of transmitted and reflected components. If the wavelength is larger than the average grain size then the scattering will not be so great but the wave will deviate from its original trajectory. Reducing the frequency of the sound wave will reduce the effect of scattering. However this will also result in a loss of sensitivity to small flaws. Absorption is a consequence of the sound energy converting to heat as particles of the material vibrate about their equilibrium. Reducing the sound frequency will reduce this effect. Once again this will reduce the sensitivity to the detection of small flaws.

For a given material and constant frequency it is possible to define an attenuation factor, α. If p_s represents an initial pressure at a source then the pressure, p, at some distance, d, from the source is:-

$$p = p_s e^{-\alpha d} \tag{2.30}$$

2.14   Beam spread and divergence

Similar to attenuation, beam spread will also cause the propagating field to lose energy with distance. For the present, as in figure 2.4, sound from a fictitious point like source (diameter $\ll 4\lambda$) is assumed and absorption is ignored. The sound now propagates in a canonical shape with its energy remaining constant. However, as the area is increasing, the average intensity reduces according to an inverse square law (proportional to $1/d^2$).

Although the average intensity reduces with distance this is not uniform over the wave's

front [24]. For any arc, centred on a line normal to the point source, the intensity reduces symmetrically with increasing angle. Beam spread is measured as the angle of the arc where the intensity is one half that at central maximum. The ratio between two measurements of power or intensity is often expressed using the decibel (dB). For example if I_0 represents the intensity value at a reference point, in this case the centre, and I_1 the intensity at another point then the ratio between the two can be expressed as:-

$$R_{dB} = 10 \log_{10}(I_1/I_0) \tag{2.31}$$

For beam spread the ratio at the half power point occurs where $I_0 = 2I_1$ so that:-

$$R_{dB} = 10 \log_{10}(0.5) = -3 \text{ dB} \tag{2.32}$$

In practice most ultrasonic transducers do not measure power or intensity directly. Rather, they produce a voltage that is proportional to the amplitude of the pressure wave. As acoustic intensity is proportional to the square of pressure amplitude, $I \alpha p^2$, the dB level is expressed in terms of acoustic pressure:-

$$R_{dB} = 10 \log_{10}(I_1/I_0) = 10 \log_{10}(P_1^2/P_0^2) \tag{2.33}$$

Alternatively:-

$$R_{dB} = 20 \log_{10}(P_1/P_0) \tag{2.34}$$

In these cases the convention is to multiply the logarithm by 20 rather than 10. For this work all measured ultrasonic signals are amplitude values. Consequently a multiplication value of 20 is applied whenever units of decibel are used. Therefore, using pressure values, the -6 dB limit is often used as a measure of beam spread. Whilst beam spread is a measure of the whole angle from side to side of the beam's centre line, beam divergence is a measure of the angle on one side only, beam spread being two times greater than beam divergence.

Beam spread occurs because vibrating particles, as considered by the one-dimensional wave equation (equation 2.9), do not just transfer energy in a single direction. For example if particles are not directly aligned to the direction of wave propagation, energy transfers at other angles. For a given application beam spread is largely determined by the transducer. Beam spread receives more attention in the next chapter which demonstrates a relationship between the dimensions of the transmitter and the wave length (λ) of the emitted sound.
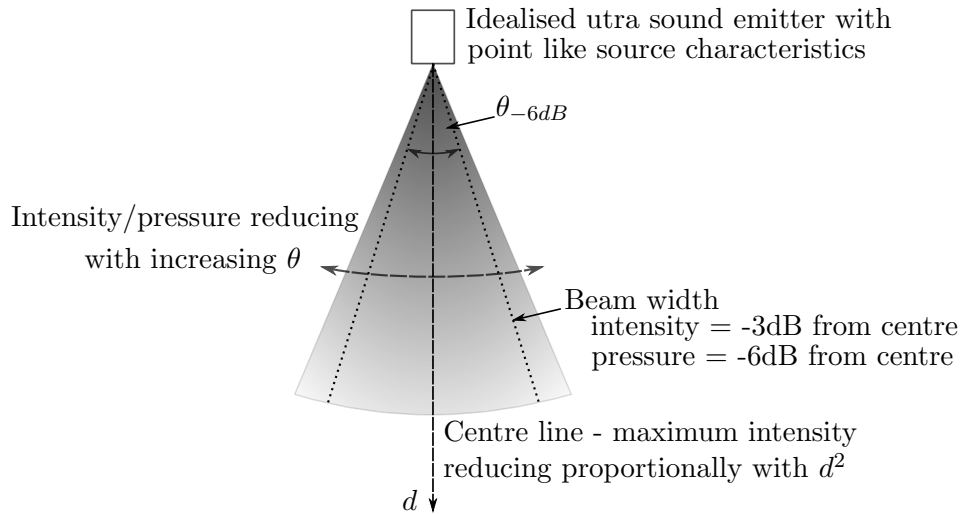
Figure 2.4: Beam spread

In particular it will be shown that if the diameter of a circular transducer is small compared with λ then the sound field is divergent. A larger transducer transmitting sound into the same media has a more directed sound field. Over large distances the transmitter with smaller diameter will have a lower inspection sensitivity, [12].

After introducing the basic concepts of ultrasound this chapter has now reached the point of discussing beam spread and propagation from a source. The next step of this introduction is to discuss the concepts of ultrasonic transducers, their operation, characteristics and methods of focusing. These subjects are covered in the following chapter which concludes the introduction to ultrasonic principles.

Chapter 3

Ultrasonic transducers, characteristics and focal laws

Following on from the introduction of ultrasound this chapter introduces the phased array transducer. The discussion includes a description of the operation of the device, its characteristics and two common methods of operation.

An ultrasonic transducer is a bi-directional device which converts an electrical signal to an ultrasonic wave and vice-versa. For practical experiments this work will use a transducer with number of small piezo-electric elements arranged in a regular linear pattern, the so called phased array. However before introducing this device it is appropriate to start with a description of the single element probe. Following this overview the chapter continues by describing the operation of the phased array. In particular it describes a method for creating a focal law allowing the acoustic beam to be focused and steered. Some attention is given to the radiation patterns and various methods of data visualisation.

3.1 The piezo-electric effect

Conventional ultrasonic transducers employ the piezo-electric effect to convert an electrical signal to a mechanical vibration. The piezo-electric effect describes the phenomenon whereby the application of a voltage across the opposite faces of a thin piezo-crystal causes its molecules to align with the electric field and deform slightly. This is reversible, so that any deformation of the material by an external force, such as a sound wave, induces a voltage across the crystal causing a current to flow into any connected circuit. There are many

naturally occurring piezo-electric materials but for ultrasonic NDT the most common material is man-made Lead Zirconate Titanite (PZT) which is both a good transmitter and receiver of mechanical vibration. The simplest form of ultrasonic transducer contains a single disc of piezo-electric material.

3.2   Single element ultrasonic probe

A cross section of a typical single element circular probe [22] is outlined in figure 3.1. The active element is a disc of piezo-electric material. The wavelength (λ) of the vibration is determined by the piezo-crystal's thickness (T) as follows [25]:-

$$\lambda = 2T \tag{3.1}$$

The wear plate protects the crystal from damage as the probe is moved along the surface of the test piece. It has a typical thickness of $\lambda/4$, [25].
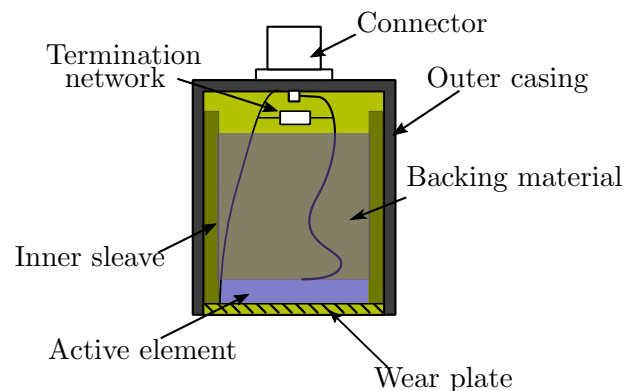
Figure 3.1: Cross section of single element circular probe

The backing material helps to damp any transient vibration. If, as is typical, the applied voltage is a short duration pulse, the transducer will respond with a highly damped sinusoid, the centre frequency and damping being determined by the crystal dimensions and the backing material. The main purpose of the backing material is, however, to reduce any prolonged oscillations to an acceptable level. In determining the damping ratio there is a compromise with sensitivity and output energy [26].

3.3 Piezo electric drive circuits

Fundamental to the operation of an ultrasonic probe is the electrical circuit responsible for initial stimulation of the crystal. For a digital electronic system the simplest way of initiating a forced vibration is by application of a short duration high voltage (typically 100 V plus) pulse. Presently this remains the most common excitation signal. Manufacturers of pulser-receivers do not publish details of their circuits and, from the user's perspective, this information is not essential. However some appreciation of the response of a piezo-crystal and current electronic system practice provides insight into the operation and set up of the pulse generator.

Typical outlines for drive circuits usually indicate a simple RC circuit with a transistor switch [27], offering no explanation for the shape of the actual pulse. To obtain a good signal-to-noise level the pulse width will typically be around $\lambda/2$, [28]. For example the ideal pulse duration for a probe with a centre-frequency of 5 MHz would be 100 nS. To transfer maximum energy, in this time, the rise time of the pulse must be fast. Using current electronic components the switching speed, and drive strength, of a nMOS transistor is approximately 2.5 times faster than that of an equivalent size pMOS [29]. Consequently it is now common for the drive voltage to be negative going. A simplified arrangement is outlined in the illustrative block diagram, figure 3.2. Many pulser-receivers provide a setting for voltage and pulse duration. The output voltage should not exceed the maximum specified by the probe manufacturer and ideally the pulse duration should be tuned to that of the probe's centre-frequency.

During reception any received signal causes a relatively small output from the piezo-element (typically a few hundred milli-volts) which must be amplified and filtered by an analogue circuit before conversion to digital. A Transmit/Receive (T/R) switch protects the low voltage data acquisition system from the high voltage excitation pulse whilst passing the low voltage echo signal.

3.4 Ultrasonic fundamentals

An assumption of the earlier discussions is that the ultrasonic waveform is from a single point source. This is unlike the face of an actual transducer which has finite dimensions. For such situations a model of more practical interest is the plane circular piston. This model considers the face of the transducer to be a circular piston oscillating within a rigid infinite baffle. The purpose of the baffle is to ensure that there is no interaction between
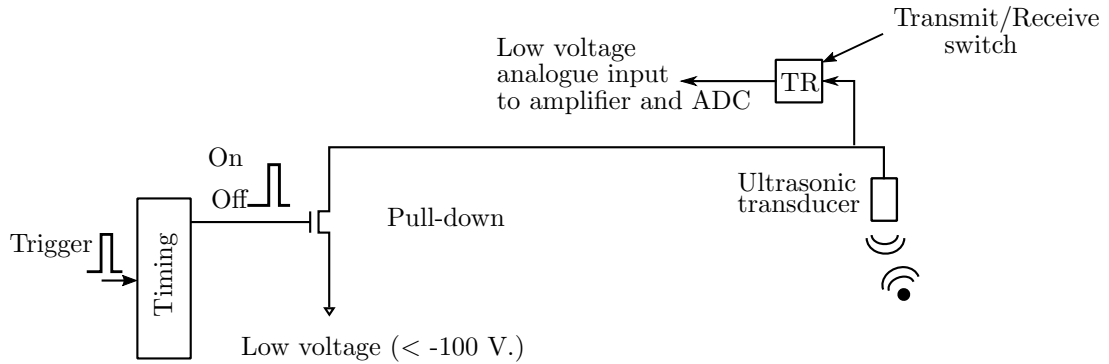
Figure 3.2: Schematic interface

displacements on the opposite faces of the piston. That is, it is possible to consider the front face waveform in isolation and without any external interference from the back face waveform.

The model further considers the front face of the piston to contain a large number of identical and minute point sources. Each source emits a spherical wave with frequency, phase and amplitude equal to that of the piston. These waves give rise to interference with several local maxima and minima making the Fresnel zone or the near field. At distances far from the sources (distances greater than the diameter of the piston's face), the wave field takes on a simpler form (known as the Fraunhoffer zone or the far field) which can be described by approximate analytic expressions.

Figure 3.3.a illustrates a piston with total surface area S and surrounded by an infinite baffle. Using this geometry and with the radiating surface of the piston moving uniformly with speed $U_0 \exp\left(j\omega t\right)$ normal to the baffle, Kinsler et al. ([30], equation 7.4.1) derive

a) Geometry for 3D analytic model   b) Geometry for 2D numerical model
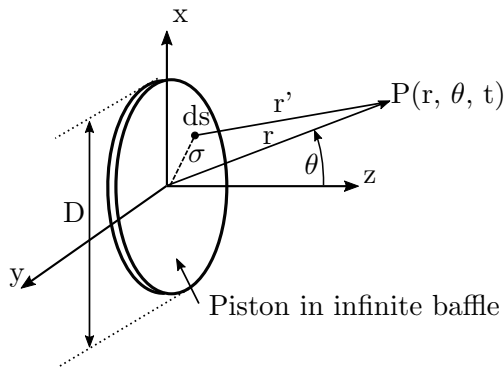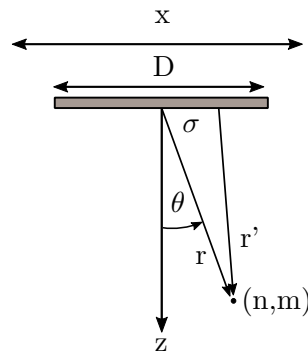
Figure 3.3: Piston model of circular probe

the following general equation to describe the three dimensional pressure field:-

$$P(r, \theta, t) \quad = \quad j\rho_0 c \frac{U_0}{\lambda} \int_S \frac{1}{r'} e^{j(\omega t - kr')} dS \tag{3.2}$$

where ρ_0 and c are, respectively, the undisturbed density and sound velocity for the transmission media.

For a general field point in three dimensional space this equation is difficult to solve. An alternative version, which describes the pressure field along an axis normal to the piston only, is equation 3.3 (from Kinsler et al. [30], equation 7.4.5):-

$$P(r, 0) = 2\rho_0 c U_0 \left| \sin\left\{ \frac{1}{2} kr \left[\sqrt{1 + (a/r)^2} - 1 \right] \right\} \right| \tag{3.3}$$

where a is the radius of the piston, is both simpler and more informative.

A resulting plot of the normalised pressure amplitude is given in figure 3.4. The axial fluctuations indicate the strong interference effects close to the surface of the piston. Pressure extremes occur where the *sin* term is fluctuating between 0 and $\pi/2$, that is:-

$$\frac{1}{2} kr \left[\sqrt{1 + (a/r)^2} - 1 \right] = \frac{m\pi}{2} \quad m = 0, 1, 2, ... \tag{3.4}$$

It is shown [30] that, moving towards the piston from large values of r/a the first peak occurs at:-

$$\frac{r}{a} = \frac{a}{\lambda} - \frac{\lambda}{4a} \tag{3.5}$$

If D is the diameter of the piston the distance to this peak from the piston's centre is:-

$$r = \frac{D^2}{4\lambda} - \frac{\lambda}{4} \tag{3.6}$$

In practice the depth of the near field is approximated to:-

$$Nearfield = \frac{D^2}{4\lambda} \tag{3.7}$$

This distance serves as a convenient demarcation between the complex interactions of the point sources and the simpler far field where the axial pressure decreases monotonically. For distances within the near field an acoustic lens, designed in a similar fashion to an optical lens, can be used to focus the sound waves. This is not possible in the far field

where the wave pattern is divergent.
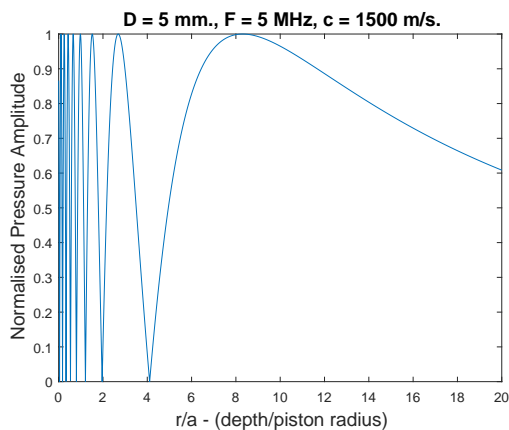
Figure 3.4: Example - axial normalised pressure distribution

3.5 Numerical simulation

Whilst an analytic solution to equation 3.2 is difficult Wooh and Shi [31] use a numerical model to create a two dimensional image of the beam profile. From Huygen's principal, which states that wave interactions can be analysed by summing the phase and amplitude contributions of a number of simple line sources, it is possible to determine the pressure at any point P(r, θ) in the (x, z) plane (figure 3.3.b) from:-

$$P(r, \theta, t) = \sum_{n=1}^{N} \sum_{m=1}^{M} P_{nm}(r, \theta, t) \tag{3.8}$$

Figure 3.5 presents the equivalent two dimensional pressure field for the example illustrated in figure 3.4 and modelled using this approach. In addition to confirming the result of the far field approximation this approach provides details of the near field and the far-field away from the central axis.

3.6 Directivity

Figure 3.5 demonstrates that the far field pressure of a transducer is not necessarily uniform in all directions. To describe this in more detail further attention needs to be given
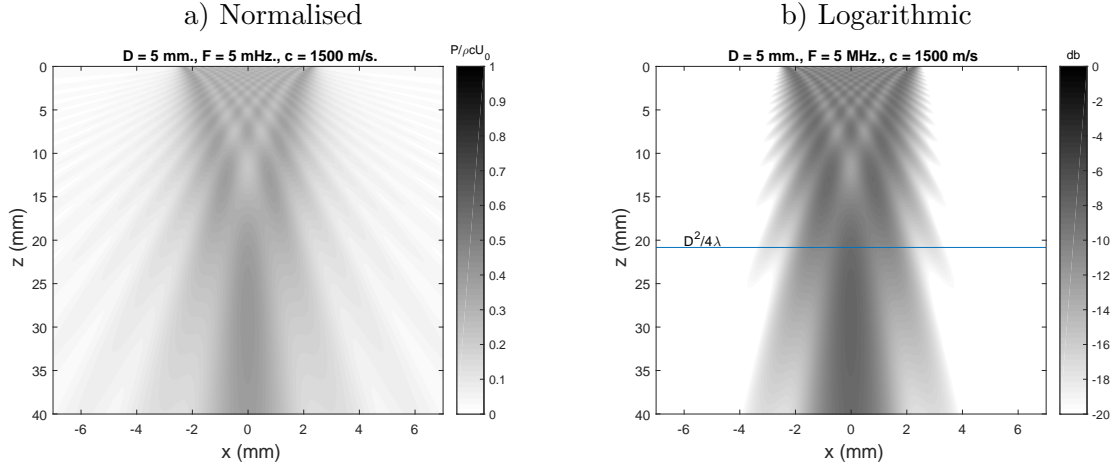
Figure 3.5: Numerical model of 2D pressure distribution

to the two dimensional model which considers the source to be a continuous line cylinder pulsating radially to generate cylindrical waves (see for example Kinsler et al. [30] or Schmerr [32]). Note that as this is a line source, as illustrated in figure 3.3.b, the approximation is representative of the two dimensional pressure field across the diameter of a circular element or across the length of a rectangular element. In the far field the pressure response of the single element is [32]:-

$$p(x,\omega) = \underbrace{\rho v u_0(\omega)\sqrt{\frac{2}{\pi j}}(kb)}_{A_p} \overbrace{\frac{sin(kbsin\theta))}{kbsin\theta}}^{D(\theta)} \underbrace{\frac{e^{jkr_0}}{\sqrt{kr_0}}}_{H(r)} \tag{3.9}$$

where the term A_p corresponds to the pressure field of a point source and $H(r)$ is the superposition of cylindrical wave terms over the entire length of the element. The term $D(\theta)$ is a directivity function representing an angle-dependent scaling of the wave field amplitude. It has the form of the sinc function:-

$$\frac{sin(x)}{x} \tag{3.10}$$

and is highly dependent on the wave number, k, and b (where b is the element length/2).

Example polar plots of the directivity function for different length elements are presented in figure 3.6. For the instance where the element length, D, is $\lambda/10$, the sound pressure is almost uniform in all directions (-90° $\leq \theta \leq$ 90°). When $D = \lambda$ there is some change in

directivity but the pressure field remains broad. At $D = 4\lambda$ the directivity is in a range of $\pm 10°$ but a number of small side lobes are now apparent. Finally with D set to 16.6λ the pattern is similar to that of figure 3.5.b which has a similar geometry and wave number.

As discussed in section 2.14 the beam spread is measured in terms of the angle at which the pressure drops by -6 dB from the value along the axis (at $\theta = 0$). For the sinc function $(\sin(x)/x)$ this value occurs at x = 1.895, therefore:-

$$\theta_{-6dB} = sin^{-1}0.6\frac{\lambda}{D} \tag{3.11}$$

The directivity of an ultrasonic element has an important role in the transducer's sound field. For the last three examples in figure 3.6 the respective -6dB angles for the main lobes are approximately $37°$, $9°$, and $2°$. In practice the elements of most single element ultrasonic transducers are typically tens of wavelengths in diameter and the sound beams are well collimated with most of the sound propagating normal to the transducer's surface. In contrast a phased array contains a number of relatively small elements and the far field directivity of single elements can vary considerably. However the transducer's overall directivity becomes a function of the number of elements used and the focal law, as discussed in the next section.

3.7 Phased arrays

An ultrasonic phased array is an arrangement of many identical piezo elements in a straight line pattern. In contrast to the introductory single element circular probe the elements of a phased array are rectangular, as illustrated in figure 3.7.

Cost is the main constraint limiting the number of elements. Not only are the probes themselves expensive but the requirement to provide each element with an individual pulse-receiver, electronic timing and physical wiring adds to the expense and complexity.

The simplest method of operation is to use each element in pulse echo mode. Firing a single element, a group of adjacent elements or all elements synthesises a variable width aperture. Alternatively elements can be fired in sequence to capture a B-scan image. Although each of these methods is likely to find some application in practice they do not take full advantage of the versatility of the phased array which lies in its ability to steer and focus the beam on both reception and emission. Cohran [33] and Olympus [34] overview various scanning methods for phased arrays. The following section introduces the basic

a) D = 0.1*λ   b) D = 1*λ

c) D = 4*λ   d) D = 16.6*λ
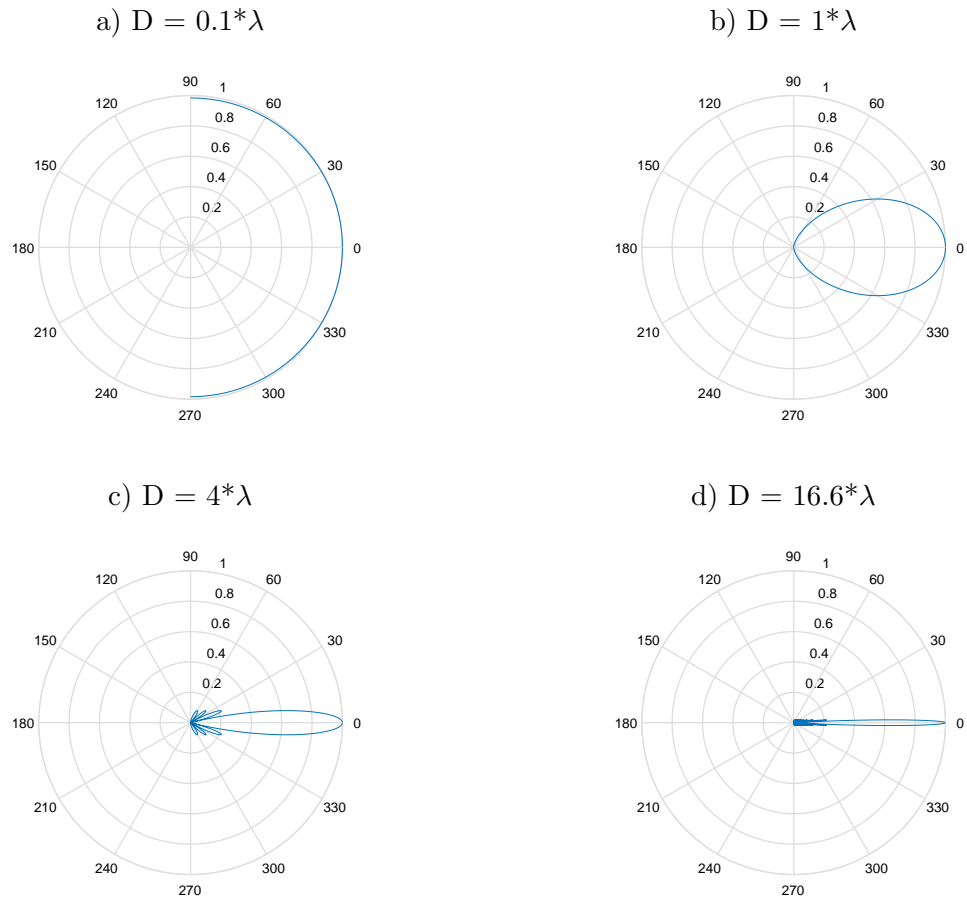
Figure 3.6: Normalised pressure at constant distance for $\lambda = 0.3$ mm
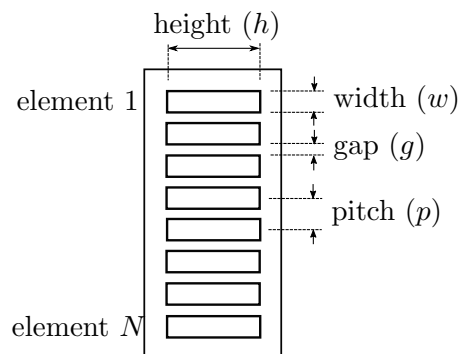
Figure 3.7: Phased array foot print

principle before developing the mathematics of beam forming and presenting simulation results.

3.8   Beam steering and focusing
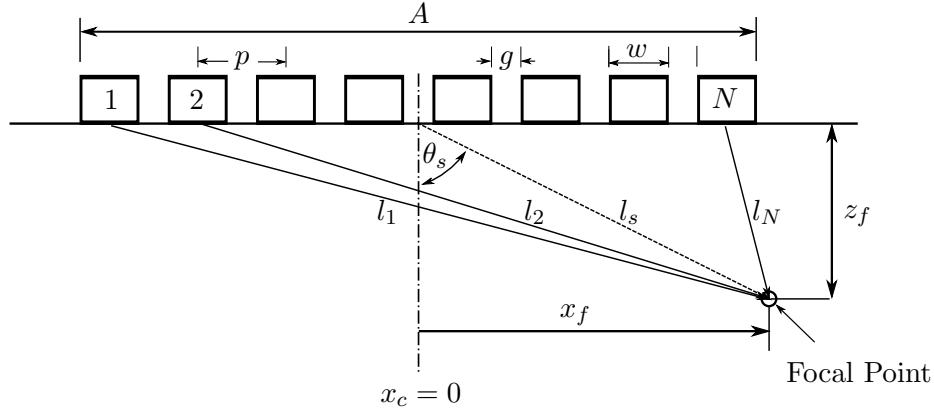
Figure 3.8: Wave paths to a focal point

Figure 3.8 illustrates the geometric paths from the centre of each element to a selected focal point. For an array with pitch p, the distance from the centre of element 1 to the centre of element i is:-

$$x_1(i) = (i - 1) * pitch \qquad (3.12)$$

Subsequently the position of each element with respect to the centre of the array where $x = 0$ is now determined by subtracting an offset, equivalent to half the distance between the centres of two outer elements:-

$$x(i) = x_1(i) - ((N - 1) * pitch)/2 \qquad (3.13)$$

Using the centre of the array, $x_c = 0$ and $z = 0$, as the reference the focal point can be expressed in cartesian, x_f, z_f or polar, $l_s \angle \theta_s$ co-ordinates. The horizontal offset x_{if} from each element to the focal point is:-

$$x_{if}(i) = x_f - x(i) \qquad (3.14)$$

and the distance $l(i)$ between element i and the focal point is:-

$$l(i) = \sqrt{z_f^2 + x_{if}(i)^2} \qquad (3.15)$$

33

For a homogeneous material the time taken for the sound wave to travel from each element to the focal point is:-

$$t(i) = \frac{l(i)}{v_s}$$

where v_s is the speed of sound in the material.

The element to be fired first (time delay $t_d = 0$) is that which has the largest distance from the focal point. The respective time delays are therefore determined by:-

$$td(i) = \frac{l(i) - l_{max}}{v_s} \tag{3.16}$$

This algorithm is readily translated into a Matlab function parameterised in terms of steering angle (θ_s) and focal distance (l_s) where:-

$$x_f = l_s \sin \theta_s \quad and \quad z_f = l_s \cos \theta_s$$

Example timing delays from this function are given in figure 3.9.

For transmission the focal law defines the time at which the excitation pulse is applied to each element. The received signal is the summation of the output of all the active elements. However before summing the output of each element is once again delayed in accordance with the focal law, figure 3.10.

3.9   Phased array directivity

The ability to apply a focal law makes the phased array considerably more flexible than the single element transducer. However the previous ray model is idealised. It does not, for example, distinguish between behaviour in the near and far field or account for beam spread and directivity. To gain further appreciation of these aspects as they pertain to phased arrays it is necessary to model the beam profile. Amongst the early and comprehensive studies in this area are the works of Wooh, Shi and Azar [31], [35], [36], [37]. Although the main objectives were to investigate the effects of various transducer parameters such as the number of elements, inter-element spacing, element width and frequency, the techniques are used here to gain further understanding of beam profiles.

The beam profile now consists of two components. Using the notation of Wooh and Shi [36] these are:-

a) Focal Distance 10 mm. in water   b) Focal Distance 20 mm. in water
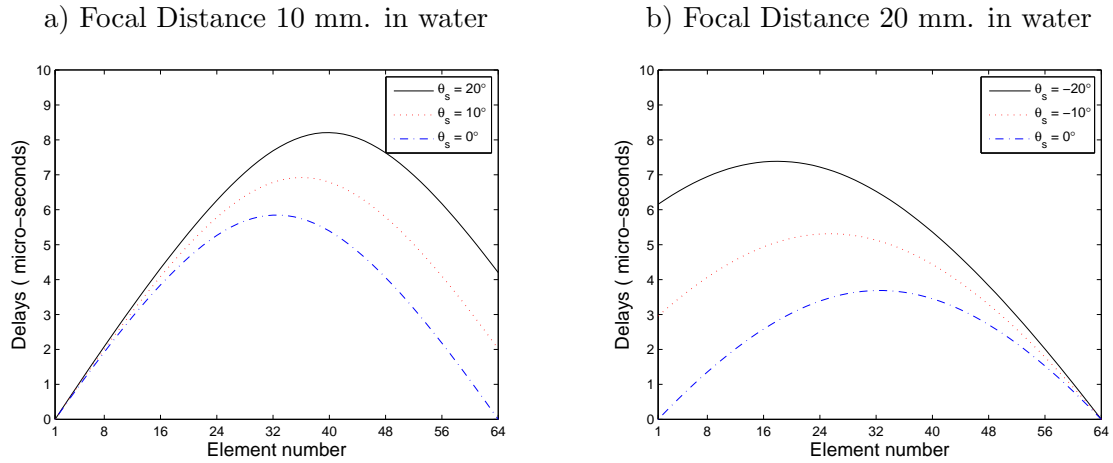
Figure 3.9: Focal law examples (64 element array, 0.5 mm. pitch)

a) Transmission   b) Reception
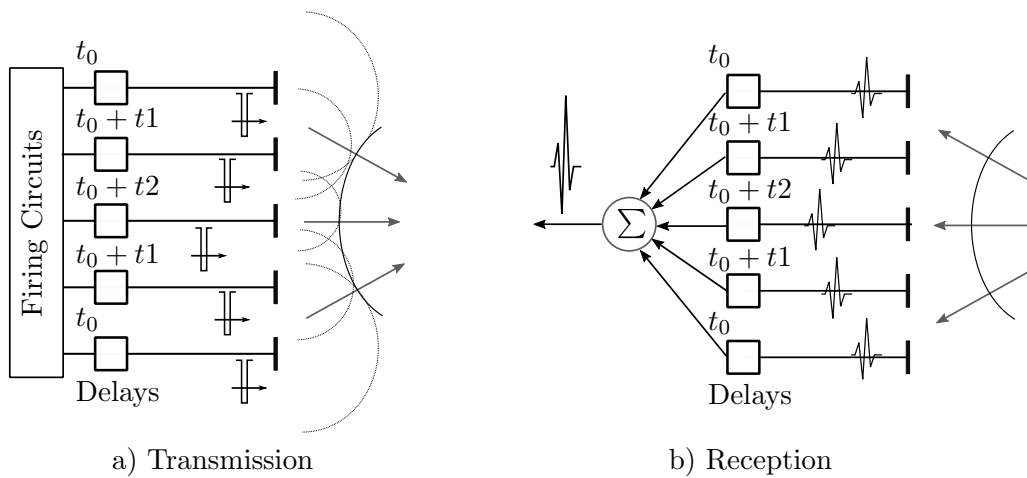
Figure 3.10: Application of focal law

1. $H_1(\theta)$ the directivity of a single element

2. $H_2(\theta)$ directivity of an array of discrete line sources

H_1 is identical to the term $D(\theta)$ in equation 3.9. In contrast to a single element probe the width of each element of a phased array is likely to be less than λ. Consequently the directivity will be relatively uniform over the range -90° $\leq \theta \leq$ 90°. This is demonstrated by the example in figure 3.11.a which illustrates the directivity of a single element with a width of 0.4λ. As the phased array contains a number of such elements directivity can be improved by pulsing all elements at the same time. This simulates a single element transducer with a larger aperture approximately equal to the sum of the number of elements and the element pitch. However as previously indicated much of the flexibility of the phased array is to steer and focus the beam through the application of a focal law. Initially the directivity of a phased array at different steering angles receives attention; a discussion of focusing is left to the section on beam imaging.

Directivity with different steering angles requires the introduction of the term $H_2(\theta)$. This term is derived by considering the phased array to be a linear array of line sources (Appendix A). From this the additional term is found to be of the form (appendix A, equation A.8):-

$$|H(x)| = \frac{1}{M} \frac{\sin(Mx)}{\sin(x)} \tag{3.17}$$

For a steering angle θ_s Wooh and Shi [31] show that for an M element phased array with pitch, p, this is equivalent to:-

$$H_2(\theta) = \frac{\sin[M(\pi p(\sin\theta_s - \sin\theta)/\lambda)]}{M\sin[\pi p(\sin\theta_s - \sin\theta)/\lambda]} \tag{3.18}$$

The overall directivity function $H(\theta)$ is the product of $H_1(\theta)$ and $H_2(\theta)$:-

$$H(\theta) = H_1(\theta).H_2(\theta) \tag{3.19}$$

The primary objective of the Wooh and Shi study is to investigate the influence of design parameters on transducer performance. In this work the available transducers are relatively limited so that the study aims to gain an understanding of the characteristics. However before considering typical transducers available for this work some results from the original paper are duplicated here. These facilitate an introduction to some of the terminology used and provide confidence in the Matlab implementation of the algorithm. Figure 3.11.b illustrates the directivity of a line source with the given dimensions. The

directivity function of the individual elements ($H_1(\theta)$ figure 3.11.a) essentially modulates $H_2(\theta)$ to produce $H(\theta)$, figure 3.11.c.

It has previously been discussed that a sharp image is produced when the acoustic energy is primarily contained within a narrow beam. Figure 3.11.c illustrates that, as expected, the maximum pressure occurs at the steering angle. However it is also observed that the main lobe is surrounded by a number of smaller lobes meaning that in addition to the main lobe ultrasonic energy is propagated into directions other than the steering angle. The smaller lobes are referred to as side lobes. These represent a leakage of energy from the main lobe but providing they remain much smaller than the main lobe the beam is well directed and the steering is pronounced.

a) $H_1(\theta)$ Single element directivity

b) $H_2(\theta)$ Array directivity, $\theta_s = 30°$

c) $H(\theta) = H_1(\theta).H_2(\theta)$, $\theta_s = 30°$.
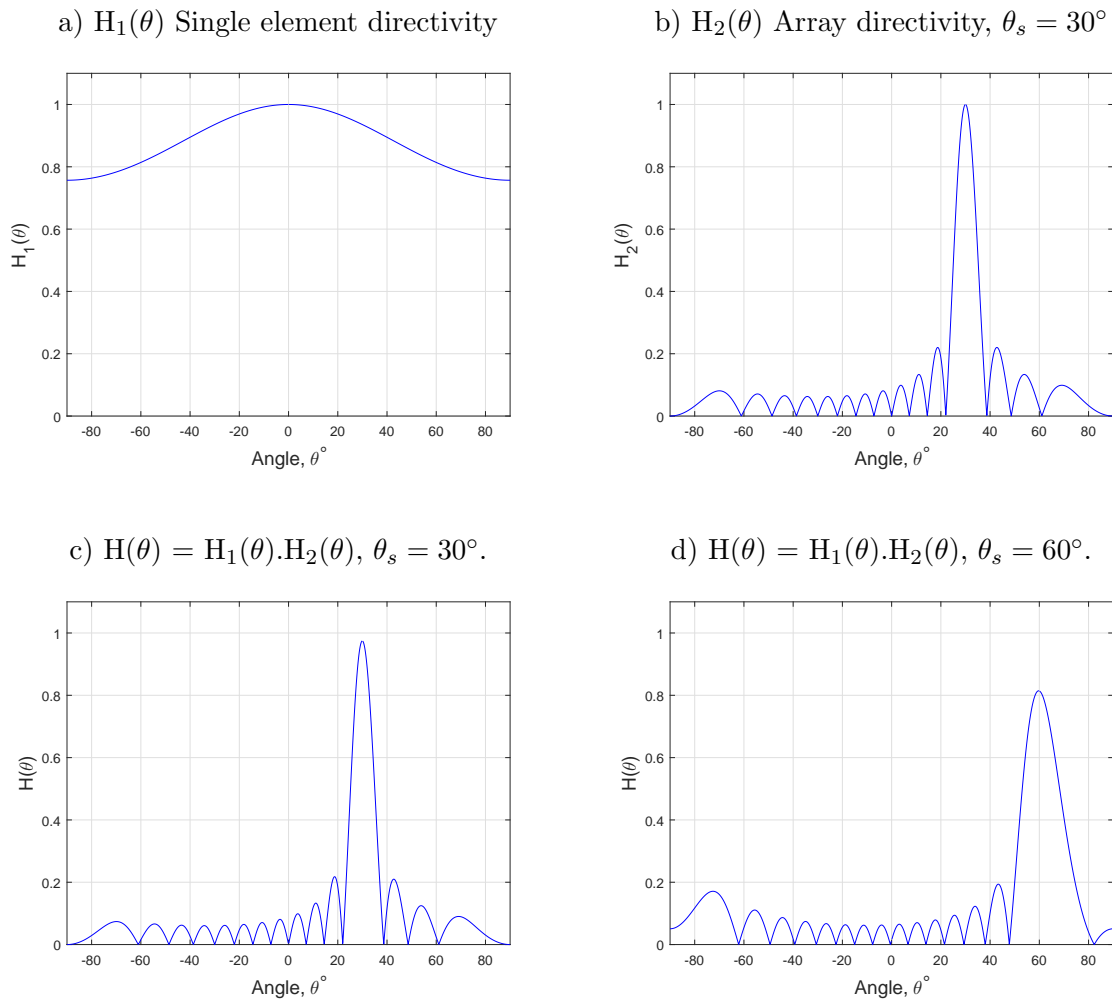
d) $H(\theta) = H_1(\theta).H_2(\theta)$, $\theta_s = 60°$.

Figure 3.11: Phased array directivity - element width $= 0.4\lambda$, pitch $= 0.5\lambda$, M $= 16$.
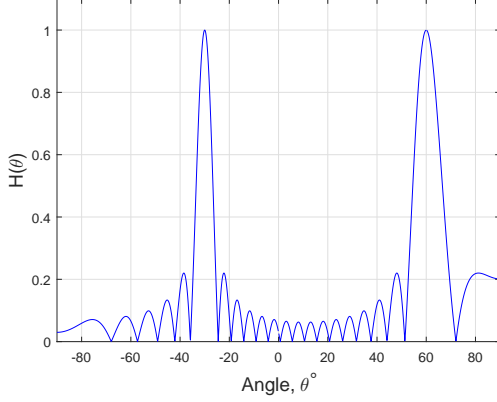
a) $H(\theta) = H_1(\theta).H_2(\theta)$, $\theta_s = 60°$.
b) Polar Plot, $\theta_s = 60°$.
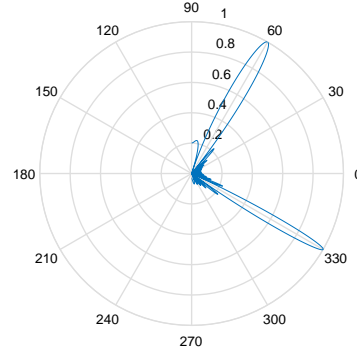
Figure 3.12: Phased array directivity - element width $= 0.01\lambda$, pitch $= 0.75\lambda$, M $= 16$.

It is also illustrated, figures 3.11.c and d that as the steering angle increases the width of the main beam also increases. More acoustic energy leaks into the side lobes at the expense of the main lobe. To complete this introductory study figure 3.12 demonstrates an extreme arrangement of element width and pitch. The example uses a small element width (0.01λ) and a relatively large pitch of 0.75λ. With a small width the directivity of each element will be uniform over -90° to 90° so that the normalised magnitude of the main lobe does not attenuate with angle. A large pitch will produce a larger overall aperture reducing the directivity of the main beam. This is demonstrated by the figure where the main beam at the steering angle of 60° has a normalised peak value of 1 and is narrower than that of figure 3.11.d. However this is accompanied by a second, narrower lobe of equal amplitude at 30°. This second lobe is termed a grating lobe. The leakage of energy in the direction of the grating lobe will return signals that lead to confusion and should therefore be eliminated.

In this example, with a large pitch, the existence of grating lobes (and side lobes) occurs at angles where the path length between elements differs by λ (or an integer multiple). They may also occur if the element width is large. Here the condition is from noting that the single element directivity function is a sinc function. Evidence of single element side lobes is present in images of figure 3.5. Returning to the phased array example equation 3.18 indicates that the maximum magnitude of $H_2(\theta_s)$ occurs when $\pi p(\sin\theta_s - \sin\theta)/\lambda = 0$. However other maxima can occur at:-

$$\pi p(\sin\theta_s - \sin\theta)/\lambda) = \pm n\pi (n = 1, 2, ...) \tag{3.20}$$

38

Solving for θ the first grating lobe (n=1) occurs at:-

$$\theta = \sin^{-1}\left(\sin\theta_s - \frac{\lambda}{p}\right) \tag{3.21}$$

If the condition:-

$$\theta = \left(\sin\theta_s - \frac{\lambda}{p}\right) < -1 \tag{3.22}$$

is met, then equation 3.21 does not have a valid real solution so that the first grating lobe does not exist in the range -90° to 90°. Consequently grating lobes can be suppressed by limiting the pitch to a maximum value of $\lambda/2$. For a maximum steering angle $\theta_{s(max)}$ without producing a grating lobe the corresponding critical pitch is:-

$$p_{cr} = \frac{\lambda}{1 + sin(\theta_s)_{max}} \tag{3.23}$$

These equations for maximum pitch and p_{cr} represent theoretical values. In practice the calculated values may need adjusting to suit the particular application.

3.10   Practical phased arrays

To enhance directivity and eliminate grating lobes a phased array may be designed with many elements and a small pitch. In practice the gap between elements is determined by the finest saw blade used to dice the element during manufacture. Around 1998 these were typically of the order 0.05 to 0.1 mm. [38], although improved manufacturing methods are likely to produce finer resolutions than the lower gap limit of 0.05 mm. At these small pitches the production of an overall aperture capable of providing a practical steering angle and focal depth requires a large number of elements. Not only is this expensive (in terms of probe manufacture, electronics and cabling) but excitation voltages are likely to unduly stress individual piezo-elements and increase crosstalk [38]. Grating lobes are, therefore inevitable.

As an example a phased array, available for this work, has a centre frequency of 10 MHz and contains 64 elements with a pitch of 0.5 mm. This particular array is for immersion applications so that, in water, $\lambda = 0.148$ mm. For $\theta_s = 0°$ figure 3.13.a illustrates that side lobes occur at multiples of $(\sin^{-1}(0.148/0.5) = 17.2°)$, the side lobe magnitude being suppressed by the modulating effect of element width ($H_1(\theta)$ term, figure 3.11).

As in the case of the single element probes it is possible to produce a two dimensional
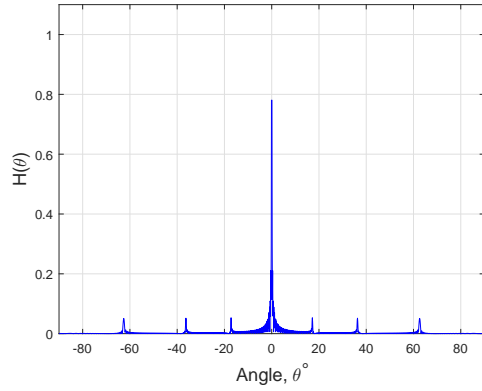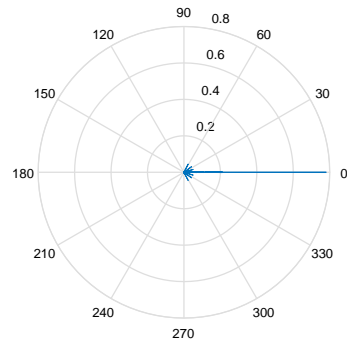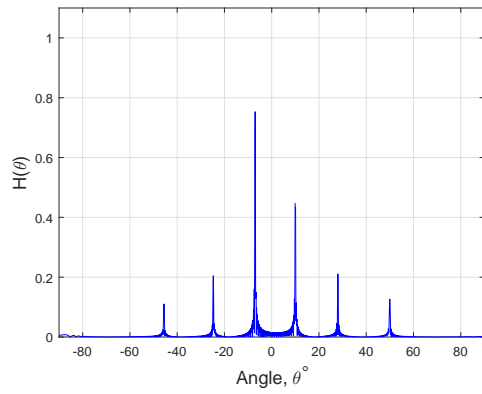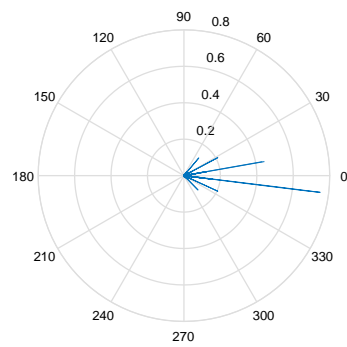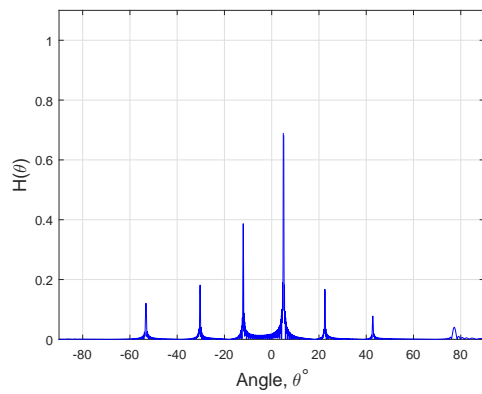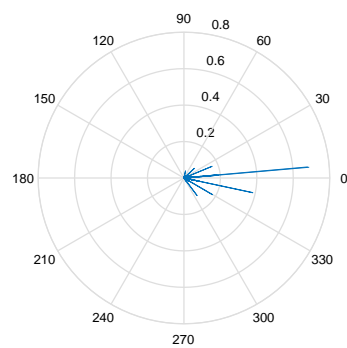
Figure 3.13: Directivity plots for 10 MHz. probe at different steering angles

pressure field using a numerical model (equation 3.8). However it is now necessary to take into account the focal law delays. Examples of the full equations, taking into account the focal law and trigonometric identities, are provided by Wooh and Shi [31] and Scmerr [32]. Using Schmerr's model, beam profiles for the 10 MHz probe are illustrated in figure 3.14. The image size is set to encompass the entire near field. With an aperture of approximately 32 mm. (elements×pitch) equation 3.7 indicates this to be approximately 1.73 metres in water.

Each figure 3.14.a, b. and c corresponds to each of the steering angles used in the previous directivity plots. In each case no focusing is applied to the phased array. Figures 3.14.d, e. and f repeat the steering angles but the applied focal law is set to focus the beam at a depth of 800 mm. In each case there is a 'thinning' of the main beam at this distance. The same applies to the grating lobe, most notable in figure 3.14.f.

3.11   Grating lobe suppression

As already indicated, and incorporated into the model, one method of suppressing grating lobes is to increase the width of the elements. By increasing the directionality a wider element (figure 3.6) improves the modulating effect of $H_1(\theta)$, figure 3.11.a. In practice other mechanisms, not accounted for in the numerical model, contribute to the reduction and in practice the $\lambda/2$ criteria (equation 3.9) can be relaxed, Drinkwater and Wilcox [39].

The first of these mechanisms is that, unlike the simulation which assumes a single frequency, phased arrays are broad band devices and excitation is by a short duration pulse. One consequence of this is that constructive interference produced by a superposition of pulses separated by a single wavelength is now less likely than that in the case of a continuous wave.

A second practical consideration, applicable to this work, occurs when a phased array is used in a dual layer set up, figure 3.15. Before outlining this operation it is first noted (as illustrated in figure 3.13.a and b) that for a single layer medium the directivity of the 10 MHz probe with no steering, $\theta_s = 0°$, has a pronounced narrow beam with relatively well suppressed side lobes. With 10° of steering there is significant degradation, figure 3.13.c and d, with the magnitude of the lobe at -7° exceeding that of the intended main lobe. At 5° as illustrated in 3.13.e and f, there is a more satisfactory response, with the main side lobe having a magnitude closer to half that of the intended main lobe.

Now consider the situation where a water medium separates the same probe from the

41
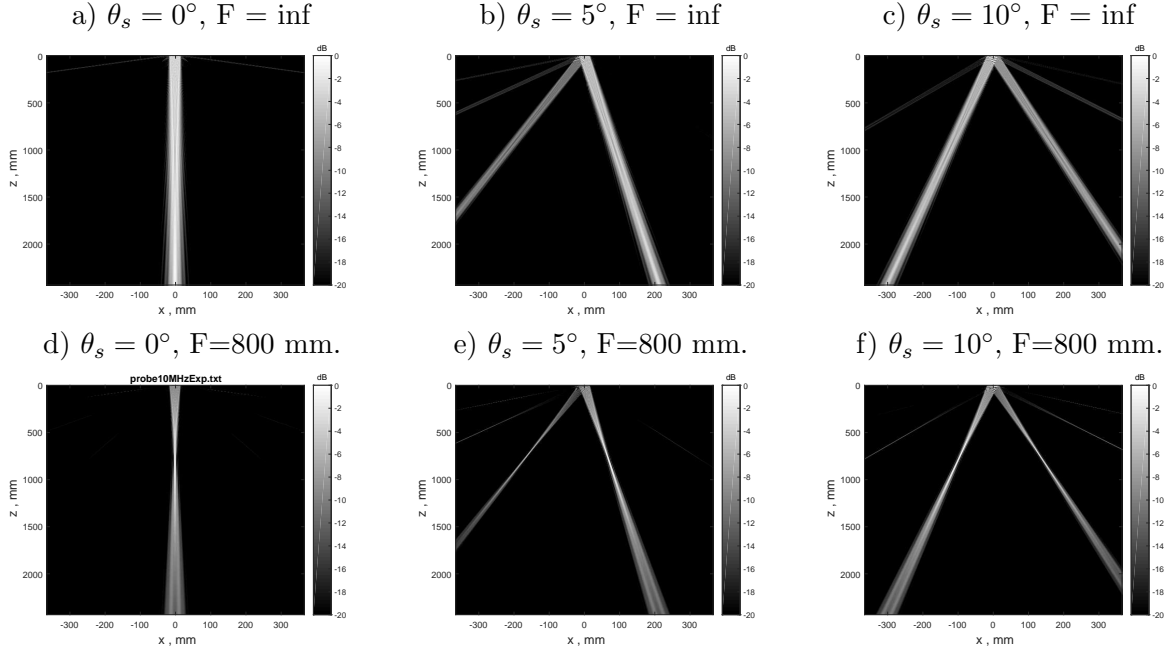
Figure 3.14: Radiation into water for 10 MHz probe at different steering angles

specimen. Steering the beam by 5° produces an angle of refraction into the specimen determined by Snell's law. Assuming approximate sound velocities of 1500 m/sec. and 6000 m/sec. for water and specimen respectively, the angle of refraction into the specimen is:-

$$\theta_r = \sin^{-1}(4\sin(5)) = 20.4°$$ (3.24)

Consequently, sweeping the main beam through the range -5° to +5° produces a sweep angle in the specimen of approximately 40°. Although a grating lobe remains present this may be weakened by further reduction in the sweep angle.

The same applies to the case of a wedged probe, figure 3.15.b. With $\theta_s = 0°$ the angle of incidence is determined by the angle of the wedge and steering is around this angle. The wedge angle is usually chosen to be beyond the first critical angle so that a predominantly shear wave is transmitted. For many applications this is a preferred set up. However, as the velocity of the shear wave is lower than that of the stronger longitudinal wave, suppression of shear wave grating lobes places further limitation on the steering angle. It is, however, likely that the direction of the grating lobes do not directly pass through the bottom of the wedge. Much of the energy is likely to reverberate within the wedge and manifest itself as a constant coherent noise at the start of the recorded time-domain signal, Crowther [40].

42

Similarly a shear wave can be transmitted by tilting an immersion probe. In this case there is no reverberation of any grating lobe. It is unlikely to reflect back to the phased array and may miss the specimen entirely.
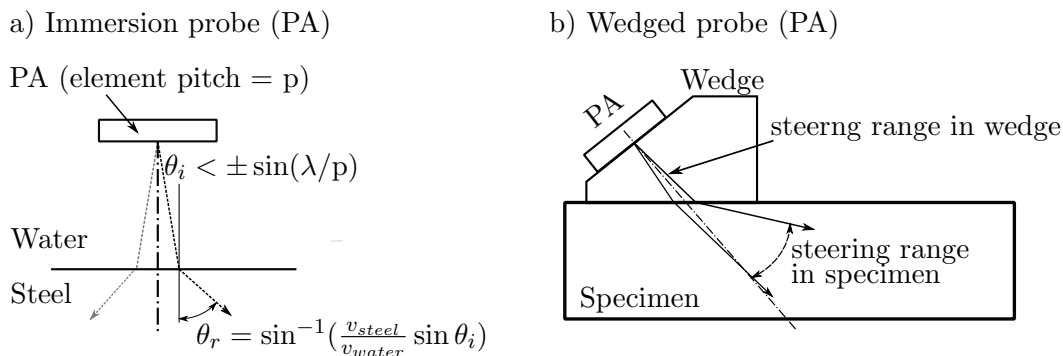
Figure 3.15: Dual layer refraction

3.12 Beam intersection point

The use of immersion probes (normal or angled) and wedges creates a complication for determining the focal law. In a typical ultrasonic test the location of the region of interest, (x_d, z_d, figure 3.16.b) is known with respect to the location of the transducer (x_s, z_s). Although Snell's law readily relates the angles of incidence and refraction, in this simple form it is of little use without first knowing the Beam Intersection Point (BIP) at x_b, z_b.

The point x_b can be determined using geometric form of Snell's law which yields a polynomial of the form:-

$$p_4 x_b^4 + p_3 x_b^3 + p_2 x_b^2 + p_1 x_b + p_0 = 0 \tag{3.25}$$

Derivation of this equation and a definition of the coefficients is given in appendix B.2

A number of techniques for solving the quadratic polynomial are available, notably Ferrari's method [41]. However in this work, the simplest and most convenient approach is to use the Matlab in-built roots function. Of the four roots x_b is given by the one, and only one, positive real number.
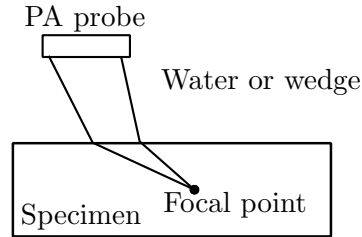
a) Physical set up   b) Refraction at planar interface
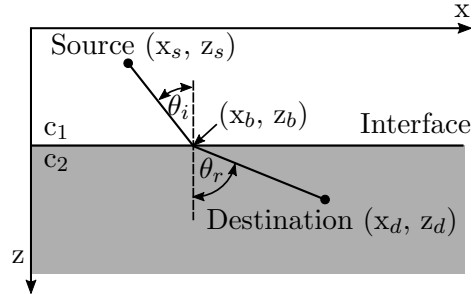
Figure 3.16: Dual layer set up

3.13 Dual layer focal law

A method of deriving a focal law for single layer media was developed in section 3.8. This method is extended to the dual layer case by first establishing the BIP for the path from the first element (1) of the aperture to the desired focal point, as in figure 3.17. Similarly the BIP for the path from the last element (N) of the aperture to the focal point is established. Using the procedure outlined in section 3.8 a focal law is calculated for the virtual focal point using the speed of sound (v_s) for the layer 1 material. A constraint is that the focal depth must be within the near field of layer 1.

The example illustrates the general case where the beam is both steered and focused. Electronic steering and focusing of the beam allows a region to be inspected without movement of the probe. In practice a more likely set up is to set the focal point at a specific depth normal to the centre of the probe. With the steering angle fixed at $0°$ the effects of grating lobes are minimised and the longitudinal wave is dominant. Alternatively the probe may be angled. Once again, without steering, grating lobes are minimised but the shear wave is likely to be dominant. Coverage of the specimen is achieved by the incremental x, y positioning of the transducer by the immersion tank's stepper motor controlled probe transport system.

3.14 A-scan

The term A-scan refers to a plot of a single ultrasonic pulse echo response. A typical example is illustrated in figure 3.18.a. In this case the transducer is the single element of a
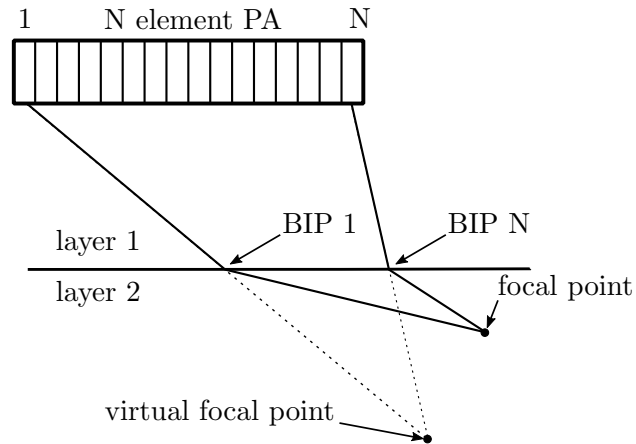
Figure 3.17: Virtual focal point across a dual layer

phased array. The initial, large, oscillations result from the front wall. The second pulse, shortly after sample number 400, is the reflection from the back wall. On reaching the front wall a percentage of energy from the back wall reflection is transmitted to the probe and recorded as the first front wall. At the front wall interface a larger proportion of the back wall reflection is reflected back into the test piece only to be reflected a second time and leading to a second back wall reflection.

If the velocity of sound in the material is known then the time delay between the start of the front wall pulse and that of the first back wall (or the time between first and second back wall reflections) can be used to gauge the thickness of the material.
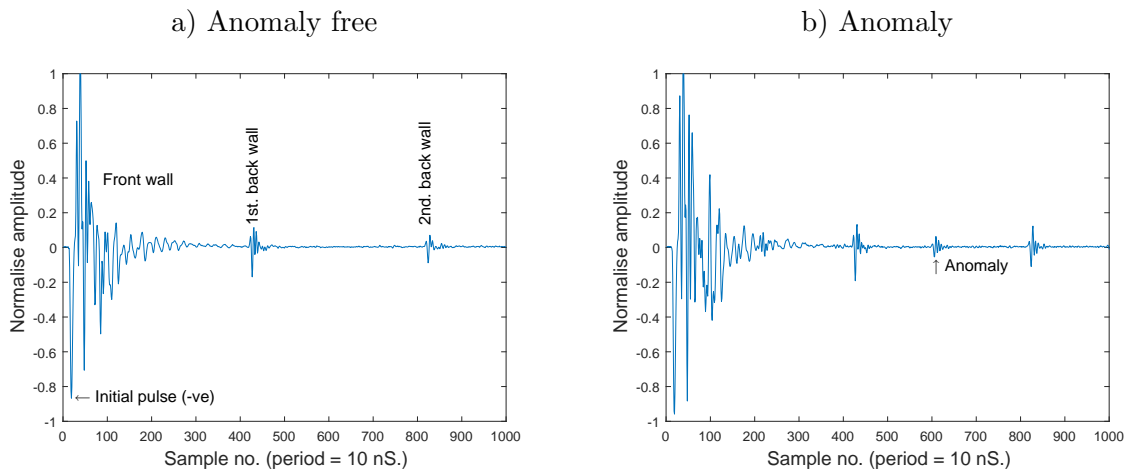
Figure 3.18: Example A-scan (contact probe)

Figure 3.18.b illustrates an A-scan from the same element after moving to a location

containing an anomaly. The depth of the anomaly can be calculated from its position with respect to the back wall reflections. Evidence of the anomaly will also be present in the front wall echo but is not so easily recognised.

The example A-scan is from a single element of a phased array. Focusing is at a fixed depth determined by the near field equation (3.7). With an element width less than a millimetre, almost all of the reflection will be from the beam's far field. Focusing the beam on a point corresponding to the centre of the test piece concentrates most of the acoustic energy at this point. This results in an enhanced response from the anomaly and a reduction in the back wall reflections. Focusing in this manner requires a number of elements to be used and application of a focal law. Through the application of a sequence of focal laws a set of A-scans can be recorded over a number of angles. This enables the creation of an image representing a cross section of test piece in the form of a sectorial scan.

3.15   Sectorial scan

An ultrasonic sectorial or S-scan is a collection of A-scans acquired by electronically sweeping a focused beam through a range of angles. An example is illustrated in figure 3.19. The focal depth of each A-scan, figure 3.20, is set to the weld's centre line so that the beam is sensitive to anomalies in the region of the weld. Recording of the A-scan does not start until a time determined by the gate. This removes unwanted front wall information and explains the curved start to the image.

Some care needs to be taken with interpretation of the y-axis on figure 3.19. It should be noted that the weld outline is actually inverted about the vertical axis. Figure 3.20 provides further information for interpreting the sector scan.

In terms of analysis this work primarily uses A-scans as a set of vectors. Images such as figure 3.19 are largely for visual information only. Details of image construction are, therefore, not described in any detail. It suffices to note that the instrument, an Olympus RDTech Tomoscan, records the data in the Digital Recording of Ultrasonic Signals (DRUS) format. In addition to the individual A-scan the, Olympus, '.rdt', file contains information from which it is possible to create the image. This includes, for example the probe offset, the individual offset of each A-scan (from which the BIP is determined), focal distance, the gate on time, gate width and wedge delays.

Further information, more pertinent to the test piece set up, is given in the following chapter.
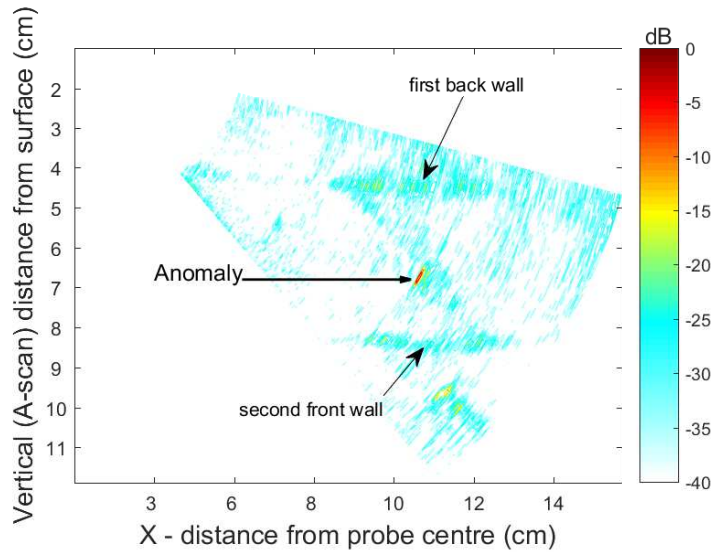
Figure 3.19: Example S-scan (contact probe)

3.16 Synthetic apertures

As a phased array contains many independent piezo-elements it is not necessary for the pulse-echo response to be restricted to that of an individual element. For example, any phased array element may record the response from a pulse on any other, single, element. Synthetic aperture focusing refers to a set of post processing techniques, that, by delaying and summing the pulse-echo responses between selected transducer elements enables the construction of a corresponding image. The Total Focusing Method (TFM) is a specific
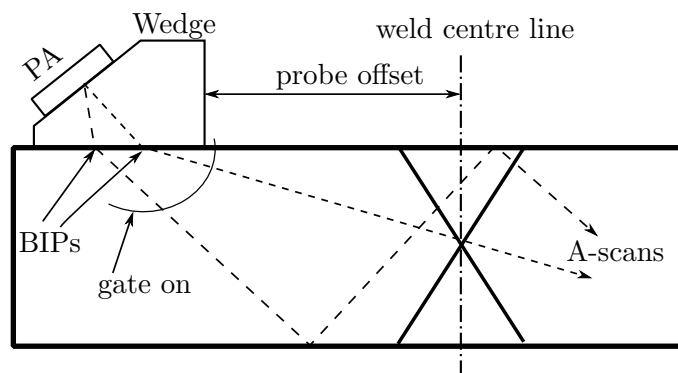
Figure 3.20: Overview of set up for S-scan

example of synthesising an image in this way and is the method adopted in this work.

To create an image using TFM (as in figure 3.21) the phased array is used in a specific manner. The most common method of phased array operation for TFM is Full Matrix Capture (FMC), an alternative is the Half Matrix Capture (HMC). A brief description of both is now given; however FMC is the method used to generate the TFM images presented later.
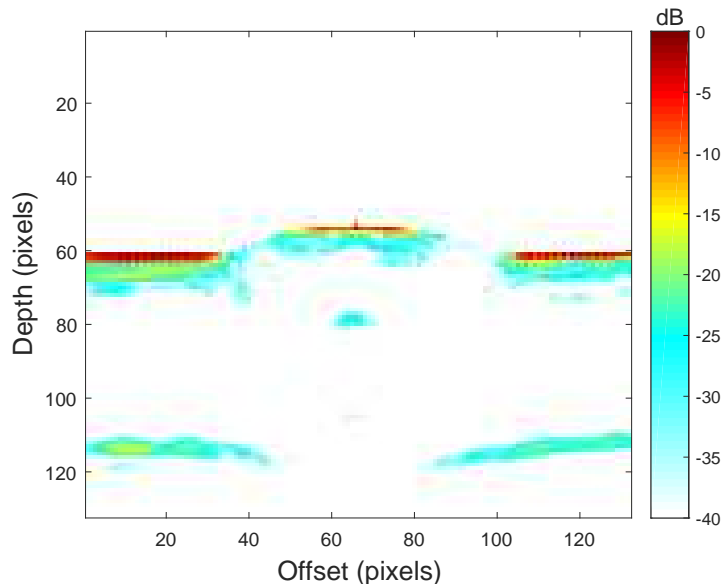
Figure 3.21: Example TFM image (immersion probe)

3.16.1 Full matrix capture

FMC does not require a focal law. Instead only a single element is pulsed at a time and all elements receive. For example, using an N element phased array the sequence starts by pulsing element 1 while all elements (including the one just pulsed) receive. This leads to the recording of N A-scans. After this, element number 2 is pulsed and the corresponding N responses are, once again, stored. The sequence repeats until every element has been pulsed and N^2 responses are recorded.

Data storage requirements and data capture time are two disadvantages of FMC. For example, if each A-scan from a 64 element probe contains 1024 samples each with a resolution of 16 bits then each FMC data set is greater than 8 MByte. With an index step of 1 mm. this results in over 8 GByte of such data for every metre of weld.

One method of reducing the acquisition time and data storage requirements is to note that through reciprocity the information content of, say, transmitting on element m, and receiving on element n, is the same as transmitting on element n and receiving on element m (i.e. the A-scan records the same path). Excluding these reciprocal paths approximately halves the acquisition time and storage requirement over FMC. The method is called Half Matrix Capture (HMC). Using HMC the number of A-scans recorded, for an N element probe, is actually:-

$$\frac{(N+1).N}{2}$$

Compared to FMC which, for a 64 element probe, captures 4096 A-scans, HMC captures only 2048. However due to, for example, the orientation of an anomaly, true reciprocity cannot be guaranteed. HMC data sets therefore contain some approximations. As suggested earlier all TFM images used in the work are from FMC data sets.

Despite its disadvantages of capture time and storage requirements, FMC has the following significant advantages:-

1. For a given transducer it contains the maximum possible data for the set up.

2. All possible focal laws can be extracted.

3. It is ideally suited to post processing and automated analysis.

For a large infrastructure project the amount of FMC data created may be vast, whilst each corresponding image has a much reduced data storage requirement. Depending on quality assurance requirements the FMC data set, after post processing, may be deleted.

3.16.2   Total focusing method

The nature of the FMC data set is such that during post processing any focal law can be applied to synthesise an image. TFM is one of a number of ways of doing this. Before describing the TFM algorithm it is useful to introduce a similar but simpler method of synthesising an image using the Synthetic Aperture Focusing Technique (SAFT) ([42], [8]). For simplicity it assumes that the probe is in direct contact with the test piece.

Both techniques start by first creating a discrete, x, z plane, into a grid. The number and size of each grid point determines the overall size and resolution of the final image. For SAFT the return flight time between an element and the selected grid point is calculated. This time is used to locate a position in the element's A-scan. The intensity value of the A-scan at this time location is now recorded. The process is repeated for the same image

point and all other elements. Summing together all intensities produces an intensity value for the selected point. If the point corresponds to an anomaly the summation results in a large intensity. Otherwise it will likely be representative of background. Although a SAFT image can be derived from an FMC data set this is not essential; it can be generated from N individual A-scans.

The Total Focusing Method (TFM) reconstruction algorithm is based on exactly the same principles, the exception being that data from every transmitter-receiver combination is used. Further, through use of the Hilbert transform, each A-scan is first converted from a scalar time series to a vector time series. This creates what is known as an analytic signal with each sample represented by a real part and an imaginary part. Every pixel of a TFM image is created using the resultant magnitude of the vector sum. Similarly a phase image may be created using the phase of the vector sum. In practice detail within a phase image tends to be more difficult to discern so that the TFM magnitude image is the more popular.

Although alternative methods of synthetic focusing continue to be investigated TFM is considered by many to be the gold standard, Bannouf et al. [43]. In addition to achieving high resolution over extended areas other authors, including Fidahoussen et al. [44] and Deutsch et al. [45] support the claim that TFM provides finer focusing and optimal spatial resolution over other methods. However for this work these claims are not of specific interest. What is more important is the fact that, particularly where data capture is automated, TFM is likely to become a commonly accepted method of image generation, the location of potential anomalies being of primary concern.

3.17   Chapter summary

Material presented in this and the previous chapter covers the basic principles upon which the ultrasonic images available for this work are based. The next chapter describes the test pieces to be used and outlines the methods of generating the ground truth for each test piece. Previous information on the organisation of sectorial data is particularly pertinent to an understanding of the interpretation of the sectorial images and the associated data capture.

Chapter 4

Test data

4.1 Introduction

All of the investigations will use the five test pieces to be described in this chapter. With each description is an overview of the experimental set up used in the particular case. Effective evaluation of a test technique requires accurate information on the location of each anomaly. Although this information is available from data sheets, in practice it is not always reliable. This is due, in part, to uncertain boundaries. In other cases there is disparity between the location of an actual reflector and that recorded in the experimental result. This may be due to movement of implants during the welding process or to misalignments with the phased array probe during test. These ambiguities and methods of establishing ground truth information will be discussed in some detail. The resulting test vectors, which define only a volume within which the anomaly is contained, are used in future chapters to automate evaluation of particular detection techniques using results from a software test bench.

4.2 Test pieces

Broadly there are two types of test piece. The first type contains a set of artificial reflectors in the form of drilled Flat Bottomed Holes (FBHs), Spherical (actually hemispherical) Bottomed Holes (SBHs) and Side Drilled Holes (SDHs). The second type of test piece contains a weld with induced faults. Each type of test piece is now described in more detail.
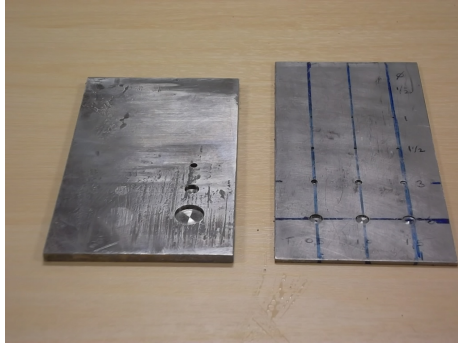
a) Flat Bottomed Holes   b) Welded Test Block

Figure 4.1: Example test pieces - a) artifical reflectors, b) induced reflectors
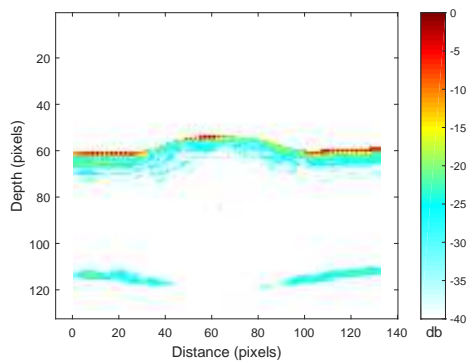
TFM Image (from TFM3)   Sector Scan (from TB1)
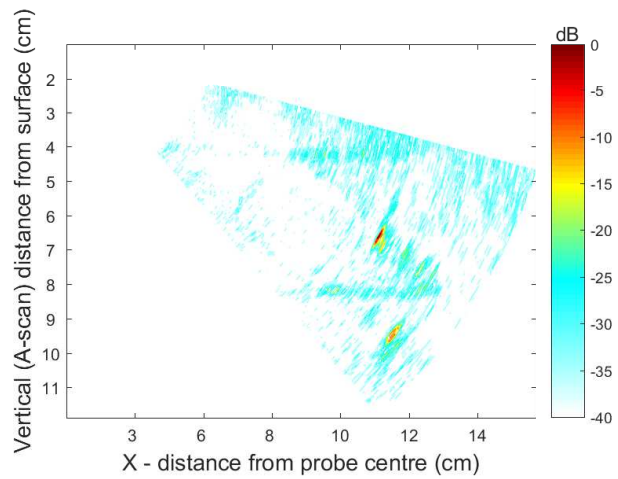
Figure 4.2: Example ultrasonic images

4.2.1 Test pieces with artificial reflectors

Three test pieces, each made from a 6 mm. thick copper-nickel material and referred to as TFM1, TFM2 and TFM3, fall into this category. Figure 4.3 provides a representative schematic for the two simplest of these, TFM1 and TFM2, each of which contains just three FBHs. The third test piece TFM3 is actually two 6 mm. plates welded together. To this, three SBHs have been drilled into the weld. The test piece also contains a SDH (20 mm. long, 1 mm. diameter (ϕ1mm)) drilled into one end of the welded test piece. Full details of the three test pieces are illustrated in figure 4.4. Neither surface of TFM3's weld is dressed.

Data for all 3 test pieces is acquired from an immersion (water) test using FMC. For this the probe is normally incident with a standoff approximately 11 mm. Each FMC data set is triggered after the probe is indexed to the next position where it remains for the duration of the acquisition. Details of the probe and set up are given in table 4.1. An example TFM image from TFM3 is illustrated in figure 4.2.a.
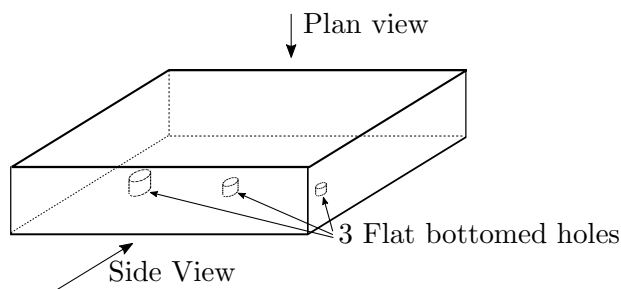
Figure 4.3: Schematic example of flat plate sample

These, high precision, manufactured test pieces provide an otherwise non-contaminated test piece. However Crutzen et al. [46] suggest that although artificial discontinuities have uses in demonstrating the capability of a technique the results may be optimistic. They further suggest that the use of artificial reflectors alone may lead to the development of an NDT technique that is hazardous when applied to structures containing real defects.

Test piece	Elements	Pitch	Length	f_c	Index step	Stand-off
TFM1, TFM2	128	0.2 mm.	3 mm.	15 MHz.	0.5 mm.	$\sim$ 11 mm.
TFM3	110	0.17 mm.	3 mm.	10 MHz.	0.2 mm.	$\sim$ 11 mm.

Table 4.1: Specification of the immersion probes
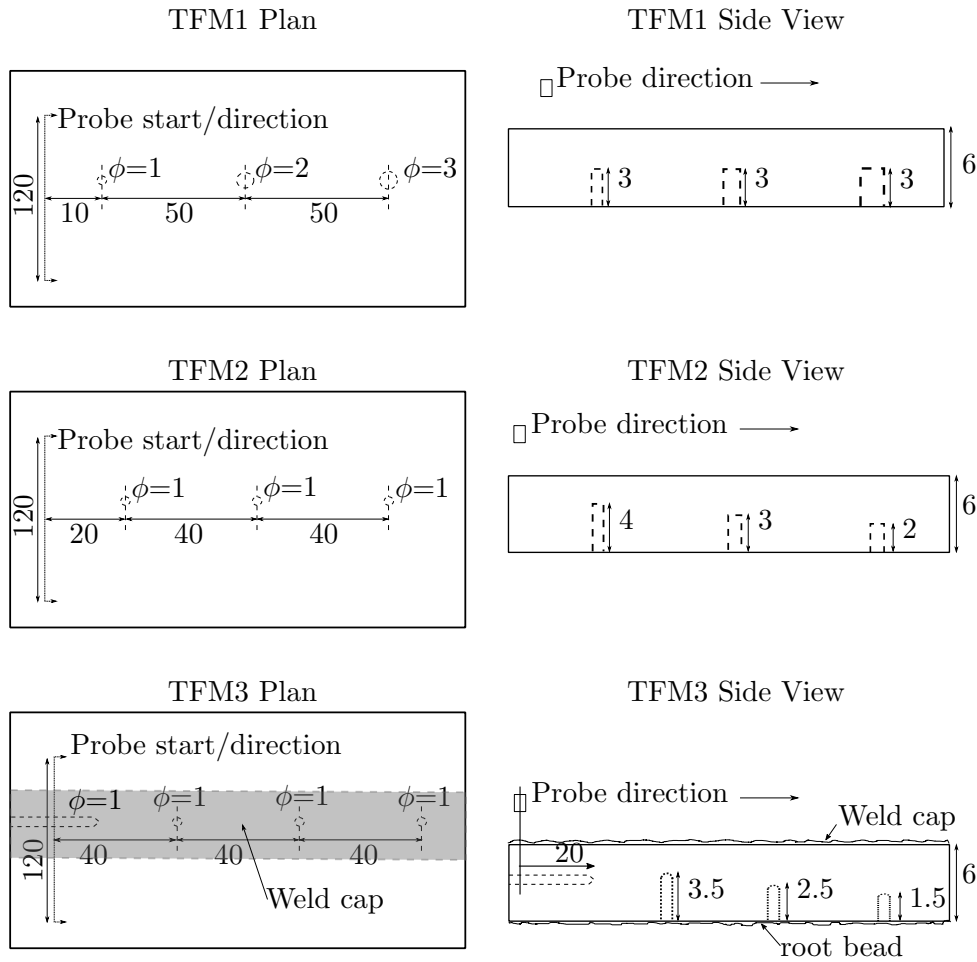
Figure 4.4: TFM test pieces (not to scale - all dimensions mm.)

4.2.2 Test pieces with induced reflectors

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. These give results that are more representative of anomalies found in a poor weld. For this work two welded test blocks, referred to as TB1 and TB2, with induced faults are used. Each contains a double-V butt weld, figure 4.6. The material is of the same thickness (44 mm.) and type of carbon steel. Various techniques are used to implant faults. For example 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 [47].

For these two test pieces data is acquired as a set of sectorial scans as the phased array

is slowly dragged along a line parallel to the centre of the weld. The phased array is mounted on a wedge which is directly coupled to the test piece using a coupling gel. Data acquisitions are triggered every 2 mm. by a string encoder attached to the phased array.

In the case of the manufactured test blocks data is available as a set of indexed sector scans as indicated in figure 4.7. Using 71 focal laws (section 3.9) the beam is scanned from $40°$ to $75°$ in steps of $0.5°$. At each angle the focal point is the weld's centre line. The figure shows that, particularly for lower angles, the beam is reflected when it reaches each wall. A reflection from the first wall is referred to as a half skip. Two consecutive skips (e.g. back wall, then front wall) is a full skip. One effect of this, figure 4.8, is that a full sweep almost covers the weld region twice. This may, in turn, cause two instances of the same anomaly to occur in a single sector scan. Of more importance is that one of two beams may receive a stronger reflection from a fault than the other. For example, consider the single fault in figure 4.8. The test piece instance of the fault is missed by the highest angle ($75°$) beam. However two other instances are likely to be present in the final image. The first reflection is captured by a lower angle beam following a half skip, the second by a yet lower angle beam following a full skip. Beam angles corresponding to the instance in the second reflection (full skip) are almost normal to the anomaly; consequently there is likely to be a strong reflection. The anomaly instance in the first reflection (half skip) is almost parallel to the beam angles. As well as presenting a smaller target there is also a greater probability that beam energy will be scattered. Arranging the inspection to cover the weld twice improves the probability of identifying a planar anomaly. This effect is likely to be less pronounced for a volumetric anomaly such as a pore.

For similar reasons it is sometimes desirable to scan the weld twice, once from each side. In figure 4.6 the phased array is first indexed in the direction of PA1. To cover the weld a second time an acquisition in the direction of PA2 can then be attempted. This improves the 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 as indicated by figure 4.6. In practice accessibility and time will likely prohibit all 4 scans, particularly if the weld is for two pipe sections. More often a maximum of two scans from a single side of the weld (e.g. PA1 and PA2) may be all that is possible.

Details of the test probe used in these inspections is given in table 4.2.

Elements	Pitch	Length	f_c	Index step
32	0.5 mm.	3 mm.	7.5 MHz.	2 mm.

Table 4.2: Specification of the contact probe

(a) TB1   (b) TB2
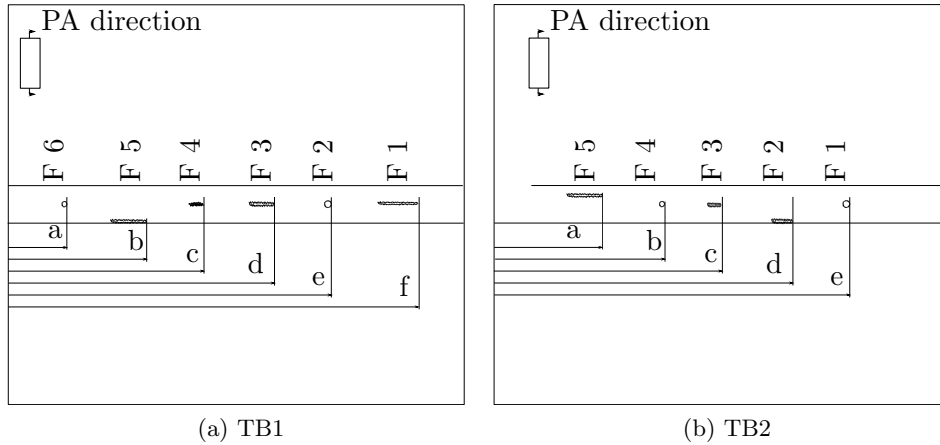
Figure 4.5: TB1 and TB2 test pieces

4.3   Data simplification

One advantage of the radiograph is that it provides an image over a length of weld. Although depth information is lost it facilitates the identification of anomalies. Ultrasonic images such as those produced by TFM or sectorial scanning provide more detail but are limited to an individual index point. For a large infrastructure project, or a continuous manufacturing process use of automated ultrasonic inspections may easily produce tens of thousands of images a day. For this work the first step in automating anomaly detec-
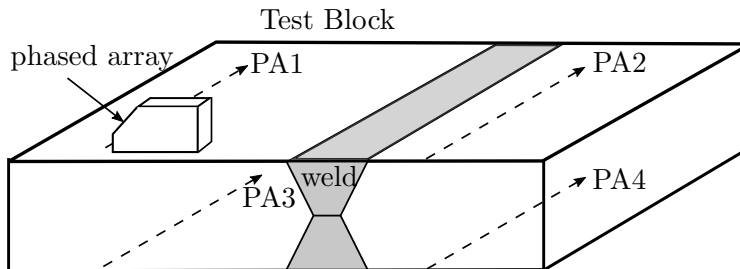
Figure 4.6: Test block data acquisitions
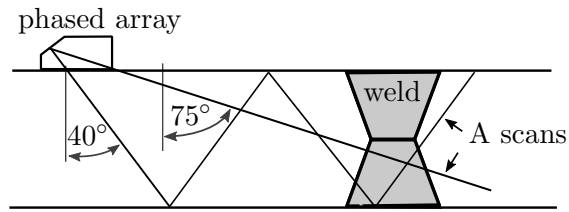
Figure 4.7: Physical scan paths
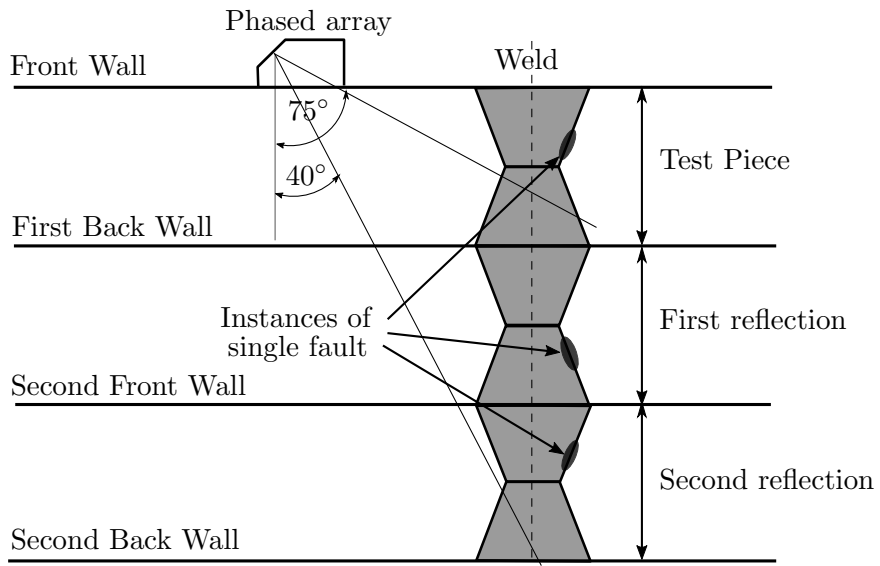
Figure 4.8: Unfolded scan paths

57

tion is to adopt an approach which, initially, models that of the radiograph. That is by compression of the set of images, representing a 3D description, into a 2D representation. Sequences of images found to contain potential anomalies can then be inspected in more detail for sentencing.

The first step of creating a 2D image is done, in essence, by taking a plan view of the test piece. Here the term plan view is used loosely. In the case of a TFM image the plan is created by extracting only the peak value from each column of the image. For the full plan this is repeated over all images. In the case of an immersion test the peak value is likely to be the first echo from the front wall revealing no meaningful information. Methods of overcoming this will be discussed later in the work.

For sector scans a plan is created by taking the peak value from each A-scan (beam angle). As the phased array is in contact with the test piece and the front wall is 'gated out' of each A-scan the front wall does not present a problem. Once again the full plan is created by repeating this exercise over a number of sectors. An example of this is given in 4.9.a. Having created this simplified view the image may be subject to interrogation by an inspector. After locating areas of interest further investigations may include interrogation of respective sets of images.

This work, however, aims to go beyond this by presenting the inspector with details of the size and location of all anomalies directly. As presented so far the obvious first step in this direction is to threshold the image. Anomalies on the resulting binary image can then be sized by, for example, counting the number of pixels associated with each anomaly. Alternatively, as illustrated in figure 4.9.b a rectangle may be drawn around each anomaly.

The example, figure 4.9, illustrates that although this simple approach does locate some anomalies there are equally other areas in the original image that might also be considered as suspicious. Reasons for this could be attributed to an inappropriate threshold level or the fact that the original image is not of a sufficiently good quality for thresholding to be effective. These are amongst the issues to be investigated in this work. However, this will require an accurate template, or ground truth, of the actual anomalies.

4.4 Ground truth data

Methods of determining a ground truth are dependent on the manufacture of a test block and the means of data capture. The following provides an overview of the methods used in each case.

a) Sectorial Scans plan view   b) Potential anomalies after segmentation
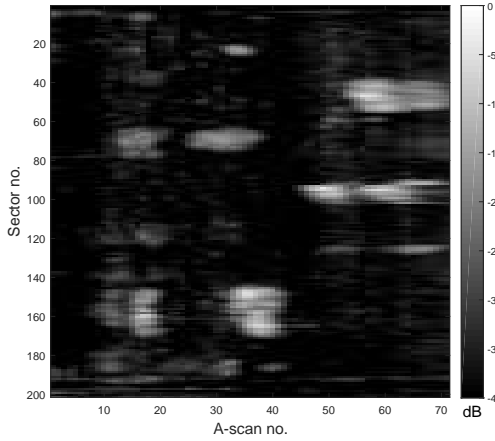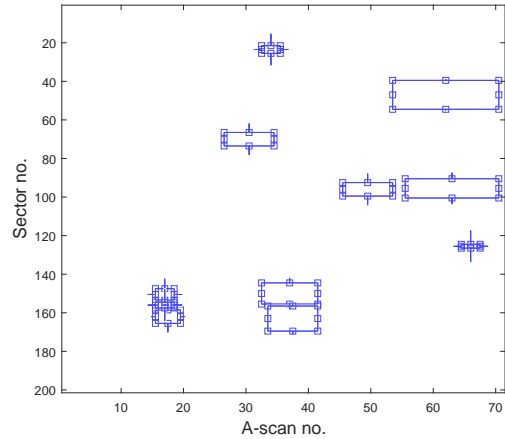
Figure 4.9: Example 'plan' view of TB1PA1

4.4.1 TFM test pieces

The TFM test pieces contain artificial reflectors making ground truth information relatively straightforward to generate from the data sheets. Examples of these, in the form of a set of masks, are presented in figure 4.10. This data is also available in numerical form which records a set of slices (image) and offsets corresponding to each anomaly. Both forms can be used by a test bench to perform automatic comparison between a derived binary image and the ground truth data.

Adding further to the accuracy of the ground truth data is the fact that data acquisition is from under immersion in a test tank. The probe is held in a mechanical manipulator and controlled, in the x, y and z direction, by three stepper motors. After adjusting the probe's height above the test piece it is then aligned with a known reflector. The manipulator, with probe, is indexed back a fixed number of steps to a start location. Now the location of all reflectors with respect to the probe's datum position are known. When ready the instrumentation system triggers a data acquisition cycle. After completion the probe is indexed to the next location and the next acquisition is triggered. Apart from some very slight miss-alignment along the line of the test piece and the direction of the manipulator the precise location of each data acquisition is known.

All three test pieces (TFM1 to TFM3) were found to contain some none intended anomalies. In most instances these are small and are ignored from the ground truth data. One exception to this is a large anomaly in TFM3. A number of investigations proved this

59

a) TFM1 ground truth
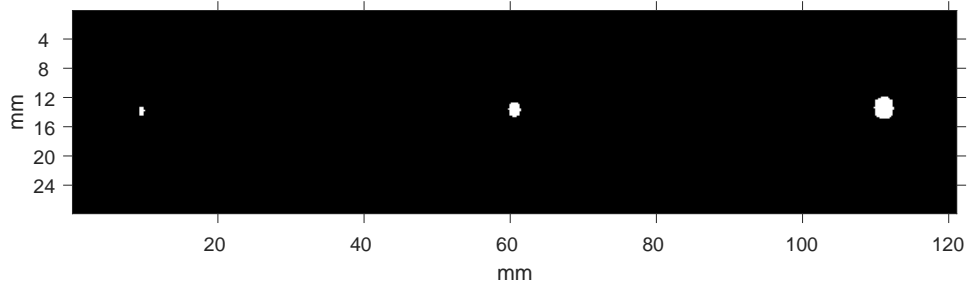
b) TFM2 ground truth
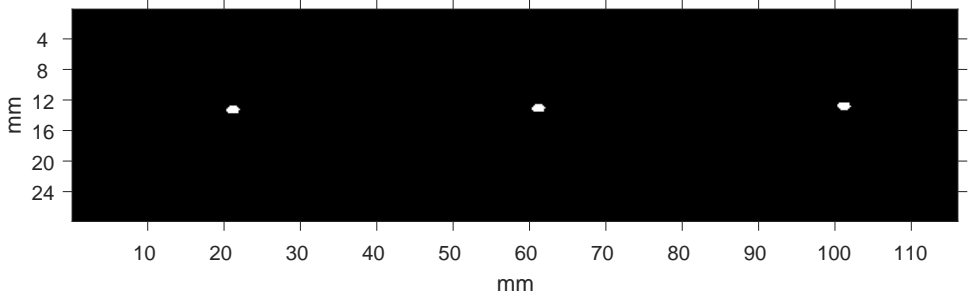
c) TFM3 ground truth

Unintended anomaly
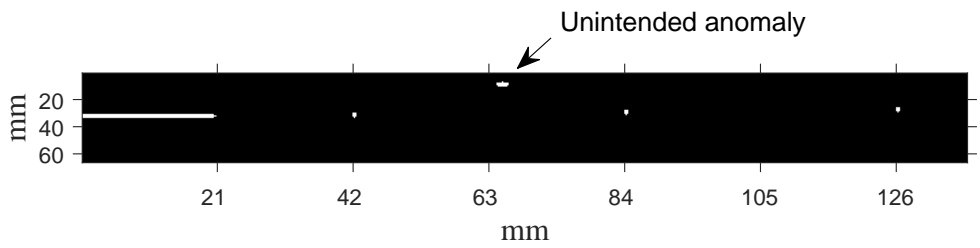
Figure 4.10: Ground truths for TFM data sets

60

to be a real anomaly rather than an equipment error or noise. As ignoring this from the ground truth would lead to false positives (indications of an anomaly where the ground truth predicts otherwise) it was included in the ground truth template. Sizing the anomaly was more difficult. Eventually and with the help of an experienced NDT inspector this was modelled as an oval with axis of approximately 1.8 mm. by 1 mm. It is located in the top half of the ground truth for TFM3 (4.10.c) around an index position of 64 mm.

4.4.2 Manufactured test blocks

Establishing ground truth data for the case of the manufactured test blocks is not as straightforward. Although a data sheet provides a good indication of the location, size and orientation of the induced faults the information is derived from the initial placement of the inserts and not an X-ray inspection after manufacture. Unlike an artificial reflector which has a geometric shape the edges of an induced reflector, after welding, are irregular and more fractal like. Movements due to the violence of the welding process, along with expansion and contraction and melting of the insert itself, all contribute to uncertainty on the precise size and location of the anomalous feature. Data sheet information on the location of each test block anomaly, prior to manufacture, is reproduced in figure 4.5. The data sheet information describes each anomaly in largely geometric terms. It does help to locate each induced anomaly and gives an indication of overall dimensions.

In the case of the test blocks an additional, and perhaps more significant, complication to generating a ground truth is due to the manner of data acquisition. In this case the probe is attached to a string encoder and manually dragged along a straight edge placed parallel to the weld, data acquisitions being triggered by the encoder. With careful manipulation the location information is accurate. However there are two caveats to this. Initially the data acquisition is triggered by the string encoder. In practice it is found that up to the first 4 or 5 acquisitions may contain invalid or poor data. More important is the fact that acquisitions stop after a fixed number of index points (in this case 217). The test blocks are 40 cm. in length and acquisitions are triggered every 2 mm. Only 200 acquisition are expected. There is now some uncertainty concerning the last few acquisitions. On inspection some of these contain distinctly invalid data, whereas in other cases the last, valid, acquisition is not clear.

To better estimate the ground truth a range of sectors for each anomaly is first estimated using the data sheet information and these are recorded. From a visual inspection of these sectors an approximate range of A-scans is also recorded. Where the same anomaly occurs

a number of times in the same range of sectors (figure 4.8) all associated ranges of A-scans are recorded. Estimation of the ground truth is now automatic. Firstly each range of sectors and A-scans is used to locate the approximate centre of the respective anomaly. A more accurate location of the anomaly's centre is determined from its centre of mass, or centroid. Figures 4.11.a to d illustrate the outcome of this for four selected anomalies (crack, inclusion, porosity and pore respectively). In practice it is found that a 25-A-scans by 25 sectors grid is sufficient to contain most of the anomaly. Finally it is necessary to establish the boundary of the region. This has been done by identifying all A-scan sector pairs that are within -12 dB of the anomaly's maximum value (for example figure 4.11.f).

Originally the region's boundary was set using -6 dB (as in figure 4.11.e), the rationale being that this value defines the beam spread and that a 6 dB drop is often used to size a feature. However this produced a ground truth where most anomalies were too small. With -12 dB there is generally a closer, but by no means precise, match between the sector size of an anomaly and that indicated by the data sheet.

Whatever method is used to generate a ground truth it is likely to be unsatisfactory in some aspect. Apart from selection of the -12 dB threshold, which is subjective, the method adopted here is consistent for all anomalies. Future tests of accuracy between test piece and ground truth will be relative, rather than attempting to give an absolute figure of merit.

The ground truth masks extracted for each data set are illustrated in figures 4.12 and 4.13.

a) Crack
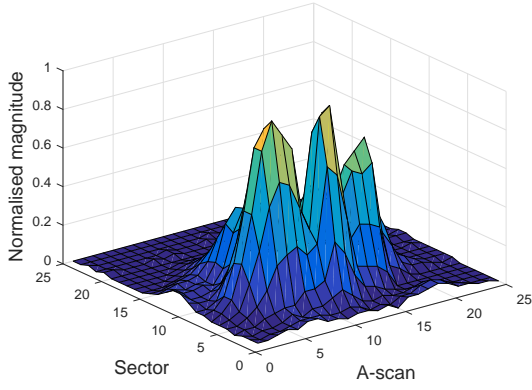
b) Inclusion
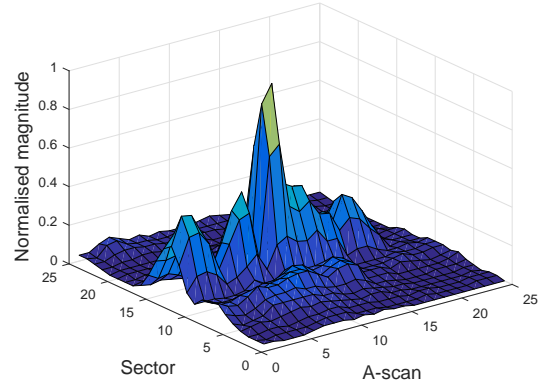
c) Porosity
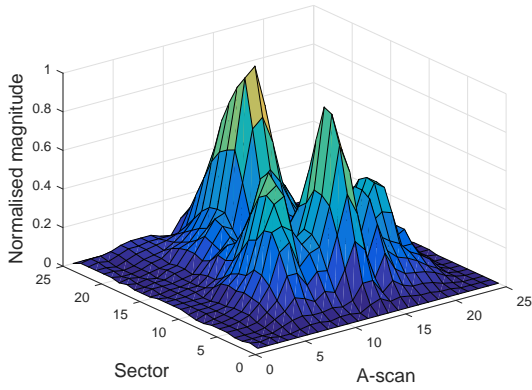
d) Pore
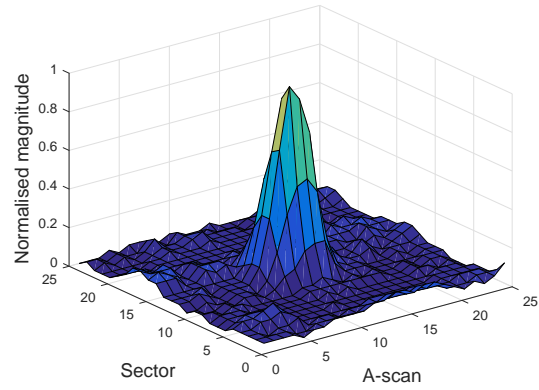
e) Crack with -6dB plane
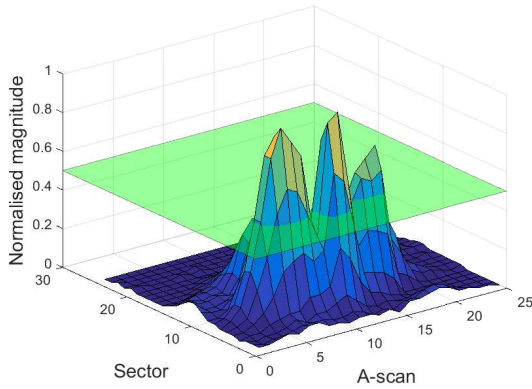
f) Crack with -12dB plane
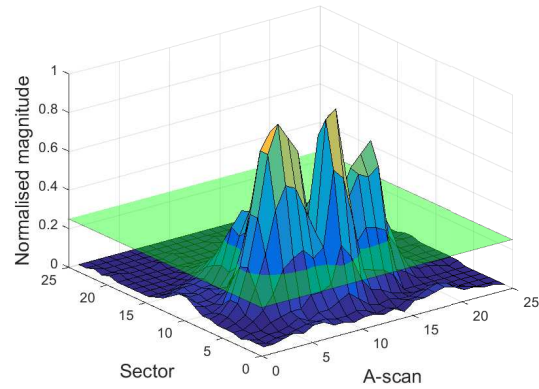
Figure 4.11: Example features and ground truth plane

a) TB1PA1
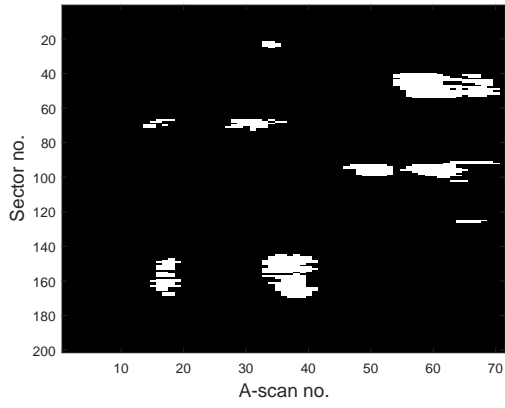
b) TB1PA2
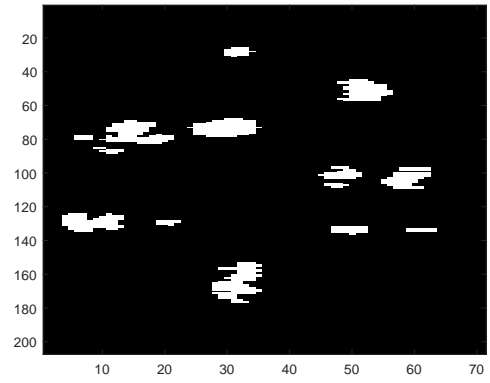
c) TB1PA3
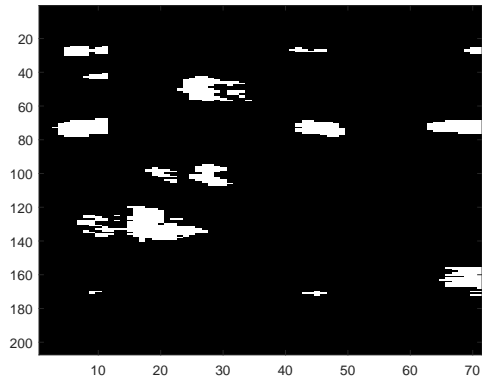
d) TB1PA4
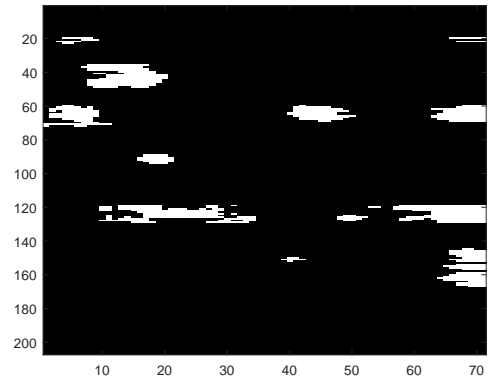
Figure 4.12: Extracted ground truths for TB1

a) TB2PA1
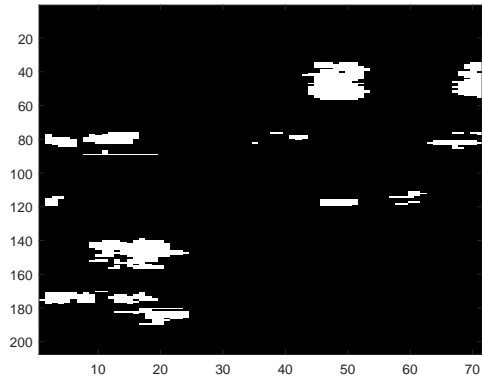
b) TB2PA2
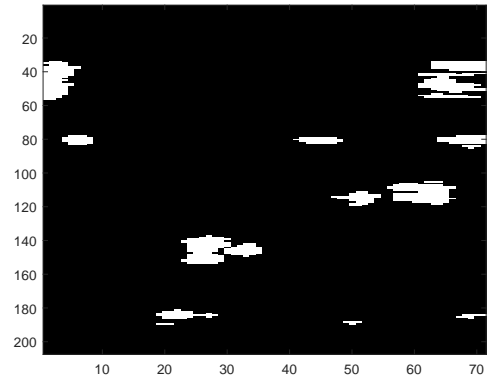
c) TB2PA3
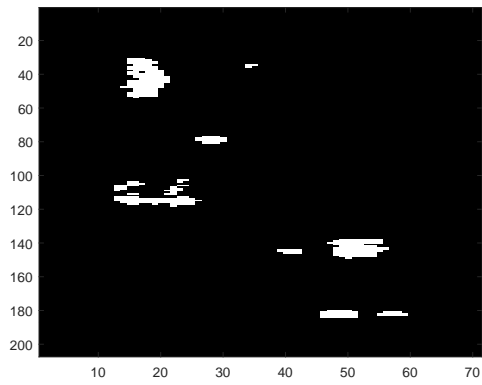
d) TB2PA4
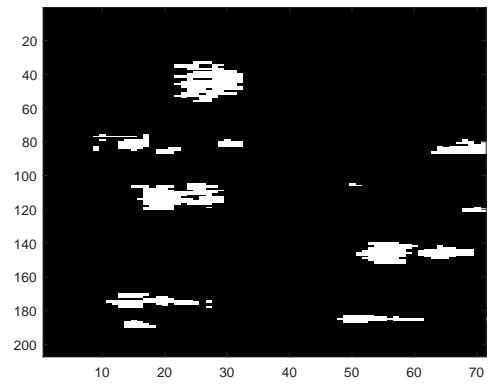
Figure 4.13: Extracted ground truths for TB2

Chapter 5

Thresholding

5.1 Introduction

An ultrasonic phased array may be used to create an image in a number of ways. The two methods used here, and previously introduced, are TFM (via FMC) and sectorial scanning. Assembling a sequence of such images at regular index points along the length of the weld provides a 3 dimensional representation of the fabrication's internal structure. Within the data set interconnected voxels containing unusually high values help to locate anomalous regions. However manipulation of this data is computationally expensive. The proposal here is to reduce the data set by taking only the peak values from a specific viewpoint, thus reducing the 3D volumetric data to a 2D image that is similar to a radiograph. Although some information is lost, correlations in the 2D data emphasise anomalous regions and the image is more amenable to efficient analysis. Through image segmentation it is possible to locate and, to some extent, size each fault. An example of this was presented in the previous chapter. In this chapter methods of image segmentation are investigated and compared.

It is immediately apparent that reducing the data set to 2 dimensions is only applicable to cases where a dominant front wall feature is not present. For example, figure 5.1.a illustrates a TFM image with front wall. Using the maximum value from each column of the image will display mainly the first front wall, masking anomalies below the weld cap. In the case of the sector scan, figure 5.1.b, which excludes the front wall, all anomalies are distinct. For this reason results presented here relate primarily to the sectorial data.

a) Example TFM Image b) Example Sector Scan
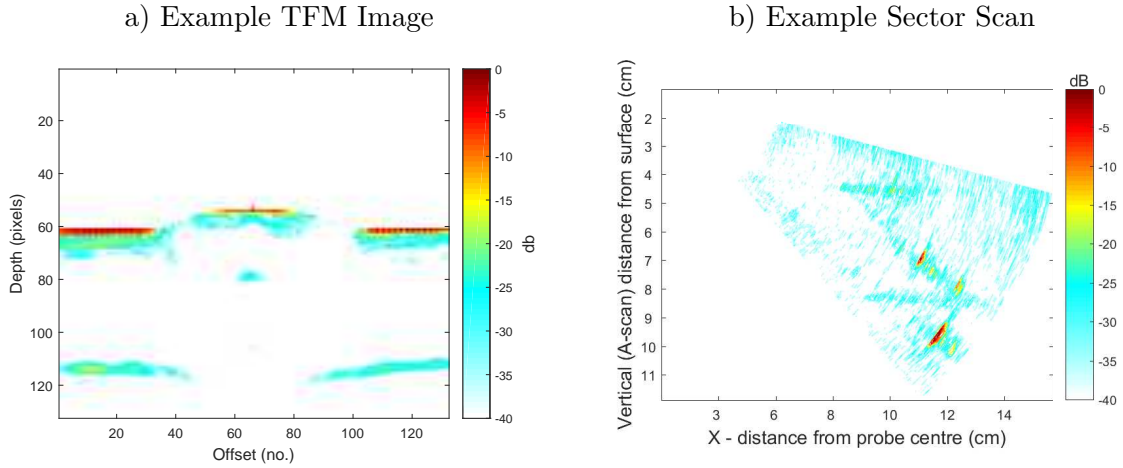
Figure 5.1: Example phased array images

5.2   Sector only anomaly detection

Before considering image segmentation in any detail notice is first given to a further simplification of the sectorial data. This involves creating a single vector in which each element contains a single peak value from each A-scan within the sector. The length of the vector is equal to the number of sectors scanned. Example plots of a vector, for TB1PA1 and TB2PA2 respectively, are illustrated in figure 5.2. These graphs are relatively straightforward to create and they give some visual indication of sectors containing potential anomalies. The location of each bar, in the form of a 'H', is taken from the ground truth test vector and indicates the position of a known anomaly. From the perspective of automating anomaly detection, a significant problem with these charts is that it is not possible to establish a reliable level whereby a set of adjacent sectors containing an anomaly are distinguished from sets that are anomaly free. In addition, according to the ground truth data, some anomaly free sectors contain A-scan peak values that are greater than those from sectors not containing an anomaly. Presenting this information to an inspector will assist with the identification of large anomalies; however smaller instances are likely to be buried within minor peaks, many of which may be representative of false alarms. The method would not be reliable and is given little further consideration in this chapter. It will, however, receive more attention in future work which discusses methods of increasing the discrimination between anomalous and none-anomalous sectors.

67

a) TB1PA1   b) TB2PA2
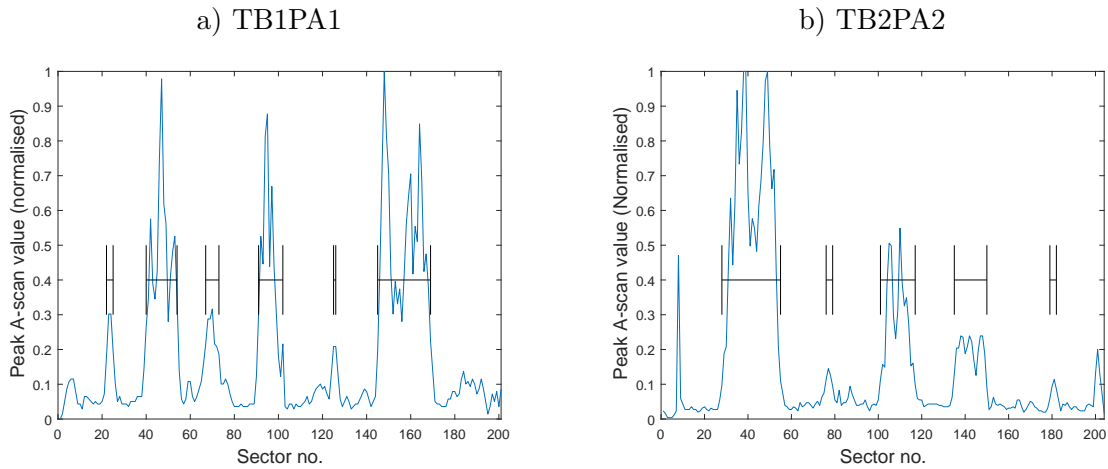
Figure 5.2: Peak A-scan value from each sector (examples)

5.3 Test block images

The set of images corresponding to the 4 scans of each test block are illustrated in figures 5.3 (TB1) and 5.4 (TB2). Evaluation of the thresholding methods is achieved by a comparison between the image after thresholding and that of the ground truth (section 4.4).

5.4 Automatic thresholding

Establishing an effective binary threshold is one of the most common operations in digital image processing. Application of the threshold is very straightforward; however estimating an optimal value is not trivial. Previous studies by Saranya et al. [48] and Sezgin et al. [49] have compared thresholding methods for Non-Destructive Test (NDT) applications. However these do not specifically relate to ultrasonic phased array data from welds.

Many thresholding methods perform well when image foreground and background areas are of sufficiently equal size and the gray levels have substantially non-overlapping distributions [50]. To a large extent this condition applies to the test blocks which, by their nature, have a relatively high density of anomalies. This is unlike many manufacturing processes where the welding operation is likely to be either by personnel with a high level of skill or automation. Under these circumstances many welds may contain large anomaly free regions. Where an anomaly does occur it might also be small.

68

a) TB1PA1

b) TB1PA2

c) TB1PA3

d) TB1PA4

Figure 5.3: TB1 images

69

a) TB2PA1

b) TB2PA2

c) TB2PA3

d) TB2PA4
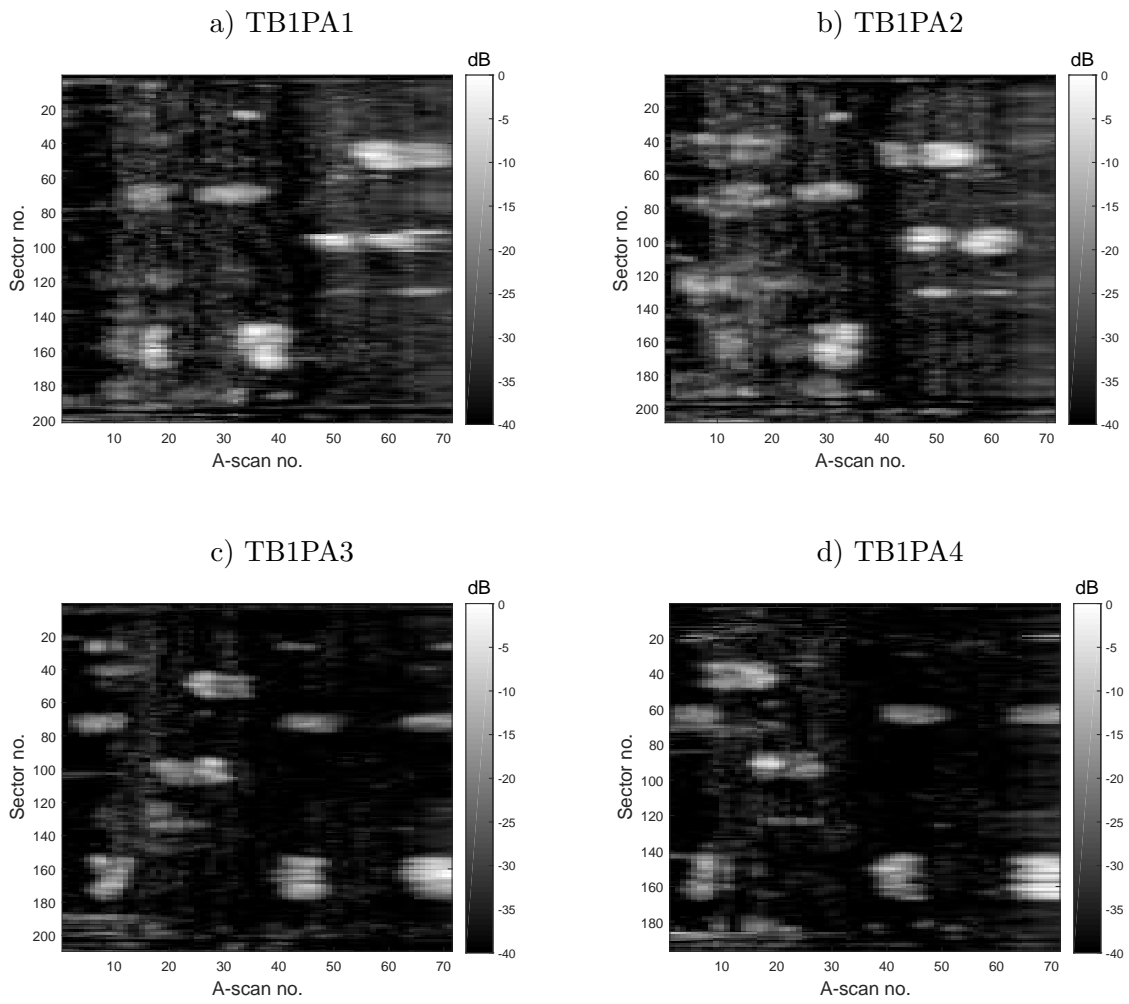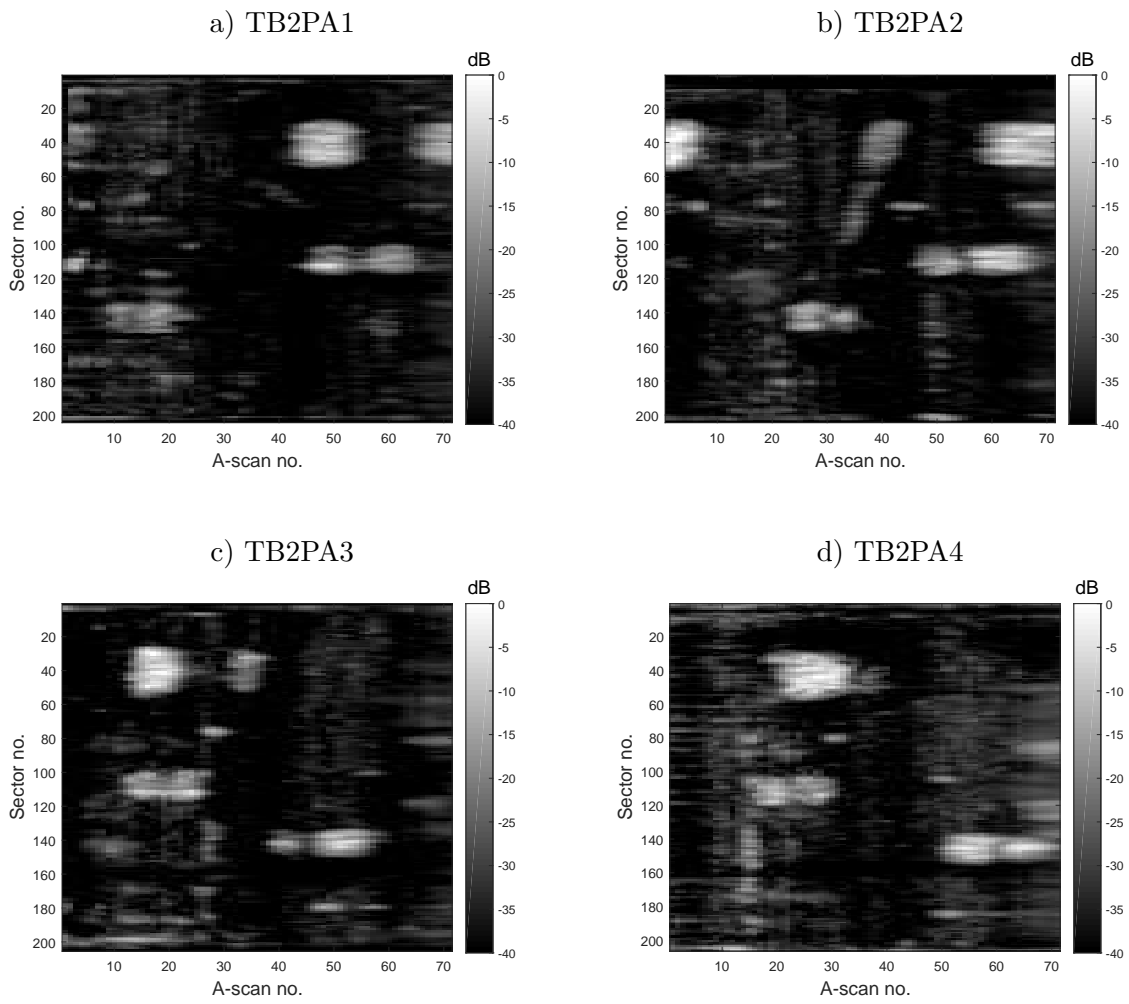
Figure 5.4: TB2 images

The following investigates three methods of automatically deriving a threshold. Each is based on the assumption that, neglecting the spatial relationships, the gray level pixel values may be treated as a random variable. A noisy background can now be modelled as a single probability density function that is different from the probability density function, or functions, of anomalous regions. Thresholding methods are, therefore based on the assumption that the gray value of pixels belonging to the foreground (anomalies) are substantially different from those belonging to the background.

Two of the methods are chosen following the work of Sezgin and Sankur [50]. Their review compares a wide range of techniques for various applications. For NDT images the Kittler and Illingworth Minimum Error [51] and Maximum Entropy [52] were recommended. To these the Otsu [53] method is included due to its popularity and availability in packages such as Matlab.

5.5   Otsu's method

Otsu thresholding [53] is based on the notion of finding a threshold that minimises the within-class variance of the resulting foreground and background classes. Consequently the optimal threshold (T) is calculated by minimising:-

$$J(T) = \frac{P_1(T)\sigma_1^2(T) + P_2(T)\sigma_2^2(T)}{\sigma^2} \qquad (5.1)$$

where, for each threshold (T) value P_1 and P_2 are the class probabilities with the threshold value set to T. The class variances are σ_1^2 and σ_2^2 whilst σ^2 is the total variance. The threshold value can be determined by iterating through all threshold values. However, by exploiting the relationship between the within-class and the between-class variances [53] it is possible to create a recursive relation that permits a faster calculation.

5.6   Kapur's method

Kapur et al. [52] consider class entropy to be a measure of class compactness and separability. They, therefore, propose a method which maximises class entropies. The criterion is:-

$$J(T) = -\sum_{g=1}^{T} \frac{h(g)}{P_1(T)} log \frac{h(g)}{P_1(T)} - \sum_{g=T+1}^{n} \frac{h(g)}{P_2(T)} log \frac{h(g)}{P_2(T)} \qquad (5.2)$$

71

where n is the maximum length of the histogram and $h(g)$ represents the histogram for the respective class. $J(T)$ is maximised to obtain maximum information (i.e $\sum_{i=1}^{n} p_i log(p_i)$ [54]) between object and background distributions. Consequently the value of T which maximises $J(T)$ is the threshold value.

5.7 Kittler and Illingworth's method

Here optimisation of the cost function is based on the Bayesian classification rule [51]. The assumption is that the bi-modal components ($i = 1|2$) of the image's gray level histogram ($h(g|i)$) have normal distributions ($\mathcal{N}(\mu_i, \sigma_i^2)$). For a given threshold T each of the two pixel populations are modelled with parameters $\mu_i(T)$, $\sigma_i(T)$ and prior probability $P_i(T)$ where:-

$$P_i(T) = \sum_{g=a}^{b} h(g)$$

$$\mu_i(T) = \frac{1}{P_i(T)} \sum_{g=a}^{b} h(g)g$$

$$\sigma_i^2 = \frac{1}{P_i(T)} \sum_{g=a}^{b} (g - \mu_i(T))^2 h(g)$$

Note:-

$a = 0$ for $i = 1$, $a = T + 1$ for $i = 2$ and

$b = T$ for $i = 1$ and $b = n$ for $i = 2$

The resulting criterion function:-

$$\begin{aligned} J(T) = 1 &+ 2[P_1(T)log\sigma_1(T) + P_2(T)log\sigma_2(T)] \\ &- 2[P_1(T)logP_1(T) + P_2(T)logP_2(T)] \end{aligned} \tag{5.3}$$

indirectly reflects the amount of overlap between the Gaussian models of background and foreground populations. As the threshold is varied models of the population distributions change. The better the fit between the data and models the smaller the overlap between the density functions and the smaller the classification error.

5.8   Threshold evaluation using original test blocks

Figure 5.5 illustrates the result of applying the threshold from each of the above methods to test piece TB1PA1. Visually it is reasonable to conclude that, in this case, the Otsu threshold provides a result that is closest to the ground truth. It may also be concluded that although all anomalies are present to some degree, in comparison to the ground truth there is some constriction of each anomaly. This leads to the danger that an anomaly may be missed completely. As an example the Maximum Entropy (ME) threshold contains no indication of fault F5, in this case the conclusion being that the threshold value is too high. Using the Kittler and Illingworth (KI) method all anomalies are present but they are dilated. In addition there is a highlighting of a number of regions not included in the ground truth.

A number of methods of providing an objective comparison between two images have been proposed, for example Zhang [55]. For binary images these largely attempt to measure the discrepancy between a segmented image and its ground truth. Examples include the Dice Coefficient (DC), the Jackard Distance (JD), Relative foreground Area Error (RAE) and Miss-Classification Error (MCE). Applications and definitions of these are found in Hubalek [56] and Sezgin [49]. All are based on a count or difference between the number of foreground pixels erroneously assigned as background (or background erroneously assigned as foreground) with respect to the ground truth image. For example the MCE is calculated from:-

$$MCE = 1 - \frac{|B_G \cap B_T| + |F_G \cap F_T|}{|B_G| + |F_G|} \tag{5.4}$$

Here F_G, B_G denote the foreground and background of the ground truth image while B_T and F_T represent the background and foreground of the image after thresholding. The notation $|\cdot|$ denotes the cardinality (or number of occurrences) of the relationship. For example if N foreground pixels of an image (F_T) correspond to N foreground pixels of the ground truth (F_G), the cardinality ($|F_G \cap F_T|$) is N.

5.9   Limitations of MCE

Table 5.1 lists the percentage MCEs (MCE $\times$ 100) for each test piece using each thresholding method. Comparison with the threshold values, expressed as a percentage of the dynamic range, indicates some consistency with the examples illustrated in figure 5.5. At face value the results indicate good thresholding. However a closer inspection highlights a

a) Ground truth   b) Otsu threshold

c) KI threshold   d) ME threshold
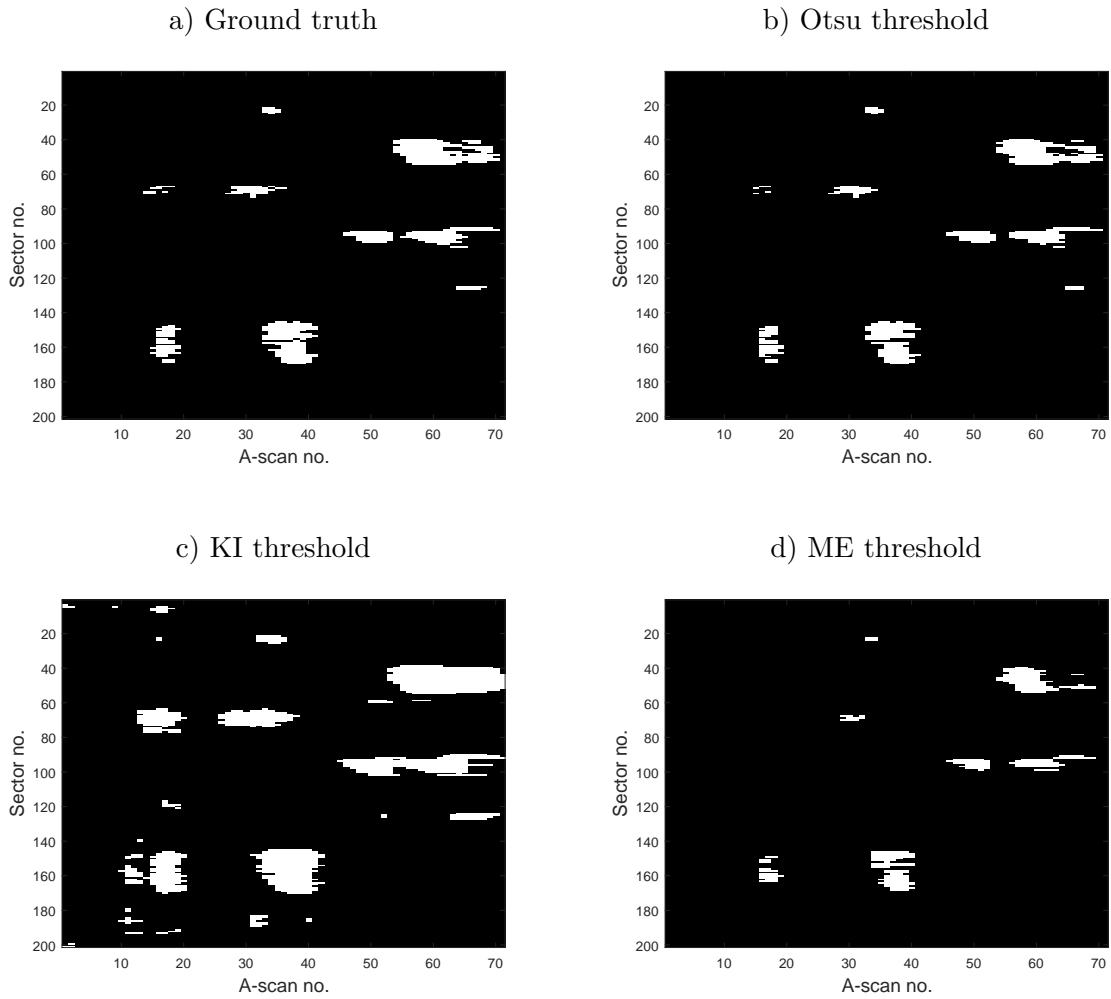
Figure 5.5: TB1PA1 threshold comparisons

problem with the MCE as a metric. Take for example the results for the Otsu and ME. The table suggests that these are closely matched. In particular for TB1PA1, the ME threshold indicates a slightly better classification than that for the Otsu threshold. However a comparison between figures 5.5.a and 5.5.d shows that the ME threshold provides no indication of one anomaly (F5) and only one instance of a second anomaly (F3). For these applications indications of all anomalies might be considered essential. In this context, preference might be given to the KI threshold. In practice the numerous false alarms will either lead to a loss in confidence or they may represent areas, that although not implanted anomalies, nonetheless represent areas containing some marginal feature requiring attention.

The MCE measure alone is, therefore, not a satisfactory method of evaluation. Consid-

Method	TB1 MCE (%)				TB2 MCE (%)			
	PA1	PA2	PA3	PA4	PA1	PA2	PA3	PA4
ME	1.3	5.7	3.3	4.8	2.5	3.4	1.5	3.0
Otsu	1.4	5.5	2.5	4.8	2.1	3.8	1.7	3.1
K&I	4.5	4.9	5.1	4.1	3.9	4.4	4.4	4.3

Table 5.1: MCEs for each threshold method

Method	TB1 - Thresh. (%)				TB2 - Thresh. (%)			
	PA1	PA2	PA3	PA4	PA1	PA2	PA3	PA4
ME	23.0	24.7	18.0	20.9	25.5	9.8	15.8	20.0
Otsu	19.4	19.1	12.9	22.2	19.5	21.6	20.2	21.4
K&I	7.6	8.3	4.1	4.6	7.7	4.9	4.7	7.2

Table 5.2: Thresholds as % of dynamic range

eration needs to be given to additional measures. This starts with a description of the confusion matrix.

5.10 The confusion matrix

When comparing a binary image against a binary ground truth each image pixel can be assigned one of four possibilities:-

- True Positive (TP) pixel is correctly identified as feature.

- False Positive (FP) pixel is incorrectly identified as feature.

- True Negative (TN) pixel is correctly identified as background.

- False Negative (FN) pixel is incorrectly identified as background.

It is common to formulate the four outcomes on a 2×2 contingency table or confusion matrix; table 5.3.

For a binary classifier the confusion matrix defines a number of rates. Some, but not all, of these are defined by table 5.3. An overall measure of the classifier's performance is the correct decision rate (alternatively known as the accuracy). This is the ratio of the total number of correct classifications to the total number of decisions:-

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.5}$$

		Ground Truth		
	Total	Positive	Negative	accuracy $= \frac{TP+TN}{TP+TN+FP+FN}$
Image — Positive		TP	FP	precision $= \frac{TP}{TP+FP}$
Image — Negative		FN	TN	NPD* $= \frac{TN}{TN+FN}$
		True Positive Rate or Recall or Sensitivity $= \frac{TP}{TP+FN}$	False Positive Rate or Fall Out or Specificity $= \frac{TN}{TN+FP}$	F$_1$ score $= \frac{2}{\frac{1}{Recall}+\frac{1}{Precision}}$

(NPD* Negative Predictive Value)

Table 5.3: Confusion matrix

This definition of accuracy has an equivalence to the MCE equation (equation 5.4) and the values in table 5.1 may be converted to accuracy by subtracting each from 100 %. The limitations of this statistic have been discussed.

It was previously suggested that despite its relatively low accuracy, the KI method has probably detected more true positives than the two alternatives, yet it has the lowest accuracy.

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

$$sensitivity = \frac{TP}{TP + FN} \tag{5.6}$$

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

Although the KI method appears to have a high sensitivity it also has many false positives, leading to many false alarms. Ideally, for detection of all true negatives, there will be no false positives. Fall out or specificity defined as:-

$$specificity = \frac{TN}{TN + FP} \tag{5.7}$$

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

Table 5.4 lists the sensitivities and specificities for each test block at each threshold level. As indicated there is a tradeoff between sensitivity and specificity. Using TB1PA1 as an example, sensitivity using the KI threshold is 1. This indicates that all true positives

are detected. However the corresponding specificity is 0.72, indicating a large number of false positives. In comparison to the Otsu and ME thresholds the indications are that the respective thresholds produce a low number of false positives. However, a number of true positives are undetected. Under these circumstances the choices are to select a threshold that gives a good probability of detection, at the expense of a relatively large number of false indications (e.g. the KI method), or to aim for fewer false positives at the expense of missing true positives (e.g. the Otsu method). In practice, if an approach gives too many false positives it may lose credibility. Alternatively in a critical situation there must be confidence that all true positives are detected.

5.11   Methods of evaluation

In the examples so far classification is by thresholding the original image. Other than selecting peak values there is no further processing. In future work thresholding remains key to the classification but different pre-processing steps are investigated. To determine whether or not an approach has any classification value it is necessary to obtain some figure of merit. An overall measure of performance is accuracy (equation 5.5) and for each threshold these are listed in table 5.4

Test Piece	Kittler Illingworth			Otsu			Maximum Entropy		
	Acc (%)	Sens.	Spec.	Acc (%)	Sens.	Spec.	Acc (%)	Sens.	Spec.
TB1PA1	95.5	1.0	0.95	98.6	0.75	0.99	98.7	0.67	0.99
TB1PA2	95.1	0.32	0.99	94.3	0.24	0.99	94.3	0.71	0.97
TB1PA3	94.9	1.0	0.95	97.5	0.50	0.99	96.7	0.27	0.99
TB1PA4	95.9	0.93	0.96	95.2	0.19	0.99	95.2	0.2	0.99
TB2PA1	96.1	0.97	0.96	97.9	0.43	0.99	97.5	0.44	1.0
TB2PA2	95.6	0.91	0.96	96.2	0.43	0.99	96.6	0.7	0.98
TB2PA3	95.6	0.96	0.96	98.3	0.49	0.99	98.5	0.61	0.99
TB2PA4	95.7	0.79	0.96	95.7	0.79	0.96	97.0	0.45	0.99

Table 5.4: Comparison of thresholding methods

The primary problem with accuracy, as a single metric, is that it suffers from the accuracy paradox [57]. It is a particular problem where there is an imbalance between the number of false positives and false negatives, [58]. As an example, consider two thresholds that produce the confusion statistics in table 5.5. Although the higher threshold produces a higher accuracy the number of false negatives is greater than in the case of the lower threshold. Consequently equation 5.5, when used alone and particularly for imbalanced

data, can be misleading. One method of addressing this is the Receiver Operating Curve (ROC), [59].

5.11.1 The ROC curve

For weld inspection, identification of all anomalies is of paramount importance. This can be at the expense of false positives but if these numbers are too large users will lose confidence in the technique. A compromise therefore needs to be made and metrics other than accuracy need to be considered.

In the case of sensitivity and specificity there is often a tradeoff between the two. As an example, and returning to table 5.5, the sensitivity figure of 0 for the higher threshold is an emphatic indication that at this level the classification is useless. Ideally both the sensitivity and specificity need to be 1.

The ROC curve [59] 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 [60]. In recent years, in recognition that accuracy is often a poor measure of performance, the approach has seen increasing use [61].

Incrementally sweeping through a range of thresholds produces a set of corresponding sensitivity and specificity pairs. At low thresholds there will be many true positives and many false positives (high sensitivity and low specificity). As the threshold increases, the number of true positives and false positives reduce (low sensitivity and high specificity). Two example ROC curves are illustrated in figure 5.6.

The method is unaffected by imbalances in the confusion statistics [62] and has two properties of value. Firstly the Area Under the Curve (AUC) provides a measure of the effectiveness of the classification. Secondly is the ability of the ROC curve to provide an

TP	FP
FN	TN

Matrix key

100	300
100	9500
Acc = 96 %	
sens. = 0.5	
spec. = 0.98	

Low threshold

0	0
200	9800
Acc = 98 %	
sens. = 0	
spec. = 1	

High threshold

Table 5.5: Example - accuracy paradox
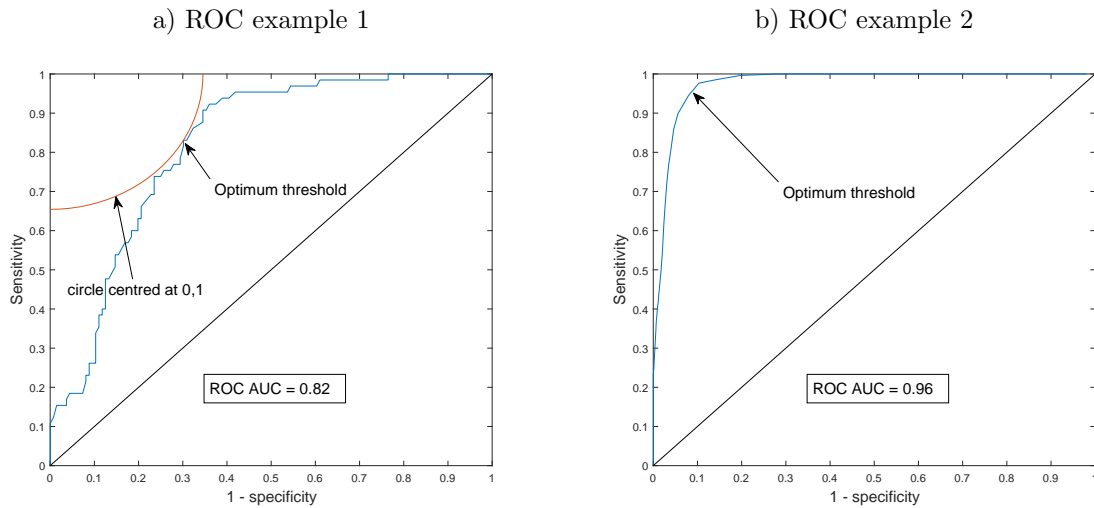
Figure 5.6: Illustrative ROC curves

optimal threshold that takes into account the costs of a trade-off between sensitivity and specificity, Greiner et al. [63]. Here the method is to select the optimal threshold as that which corresponds to the point closest to the curve's top left hand corner.

The ROC curve has two main uses. Firstly it provides an 'optimum' value, that closest to the (0,1) point, against which the threshold actually used may be compared. Secondly it gives an indication of the effectiveness of the classifier. Consider the two examples, figure 5.6. 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 5.6 lists a set of statistics obtained from ROC curves for TB1. In each case the AUC indicates good classification. A comparison of the ROC thresholds with those listed in table 5.2 indicates a closer correspondence to those obtained using the KI method.

In the following work the ROC curve is the basis of most evaluations. However for completion two other metrics are given mention. These are the F1 score and the Matthews Correlation Coefficient (MCC).

	ROC statistics					
Test Piece	AUC	Th. (%)	ACC	sens.	spec.	F1
TB1PA1	0.99	7.01	95.5	1.0	0.95	0.61
TB1PA2	0.95	4.16	90.3	0.98	0.90	0.60
TB1PA3	0.95	4.21	95.5	0.99	0.95	0.65
TB1PA4	0.93	2.93	93.0	0.98	0.93	0.62
TB2PA1	0.95	7.88	96.7	0.97	0.97	0.72
TB2PA2	0.93	3.03	91.9	0.98	0.92	0.58
TB2PA3	0.93	4.31	95.6	0.96	0.97	0.57
TB2PA4	0.89	3.28	90.1	0.97	0.90	0.49

Table 5.6: Optimum classification statistics from ROC curves

5.11.2 The F1 score

An alternative to measures based on sensitivity and specificity is that based on sensitivity and precision. These metrics have found widespread application for assessing classification performance in information retrieval systems [61]. A characteristic of these applications is that the number of true negatives in a retrieval search may be extremely high. In this context the literature tends to use the term 'recall' in place of sensitivity.

The F1 score provides a harmonic mean of precision and recall:-

$$\frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} \tag{5.8}$$

5.11.3 The Matthews correlation coefficient

This measure can be used even if the classes are of very different sizes. It has a range of -1 to 1 where -1 represents a completely wrong classifier and 1 indicates a completely correct classification. It is defined as:-

$$\frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP) * (FN + TN) * (FP + TN) * (TP + FN)}} \tag{5.9}$$

5.11.4 F1 and MCC result comparison

To gain insight into the F1 score and the MCC, figure 5.7 compares each against the full range of thresholds for TB1PA1. Both curves have a similar shape and produce very similar confusion statistics, table 5.7. Compared to the optimum threshold suggested by the ROC

curve the threshold for a peak F1 and MCC is, in all cases, higher. This reflects in the sensitivity and specificity figures which are, respectively, lower and higher than the ROC equivalent.

In terms of providing an optimum reference for evaluating the performance of a classifier these measures are, in this application, not appropriate; in particular, the lower sensitivity values indicate a greater probability of the threshold to miss anomalies. The measures may, however, provide additional information for comparing the results of a number of techniques. This would, for example, be in terms of a comparison between F1 scores (or MCC scores) after thresholding. Of note is the fact that whereas the ROC thresholds are closer to those determined from KI thresholding, the thresholds producing maximum F1 and MCC scores more closely match those of the Otsu and ME methods.

The ROC curve suggests an optimum threshold based on a tradeoff between sensitivity and specificity. It also provides, through the AUC, a measure of classifier performance. This will, therefore, be the preferred method of evaluation.

a) F1 score   b) Matthews correlation coefficient
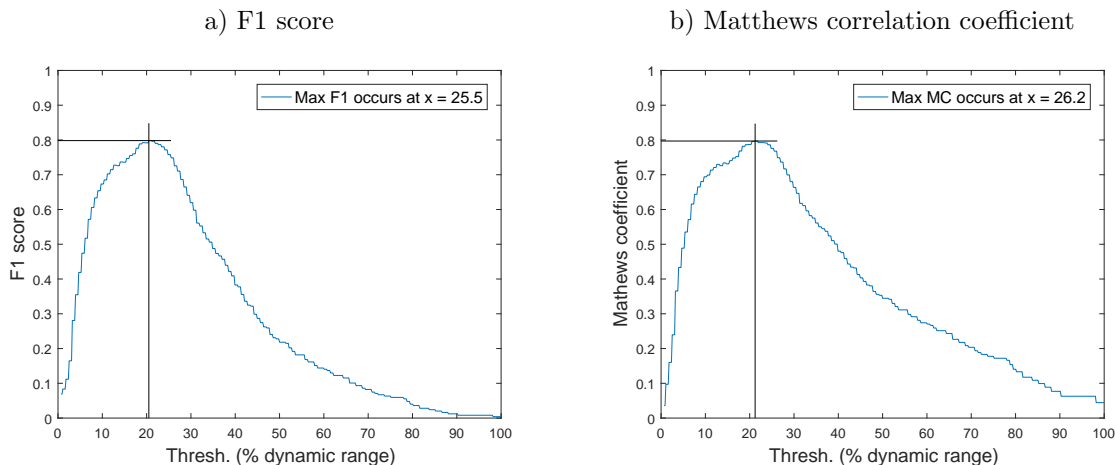
Figure 5.7: Example F1 scores and Matthew coefficients vs %Thres.(TB1PA1)

5.12   Performance measure

Investigations throughout this work will include methods of pre-processing data before classification. In many cases the classification is likely to be followed by further thresholding to investigate the possibility of improving performance. For consistency, performance measures are based on ROC analysis. In essence this requires recording figures for actual

Test Piece	F1					MCC				
	Max.	Th.%	Acc.%	sens.	spec.	Max.	Th.%	Acc.%	sens.	spec.
TB1PA1	0.79	25.5	98.7	0.73	0.99	0.79	26.2	98.8	0.71	0.99
TB1PA2	0.71	11.5	94.7	0.86	0.95	0.69	11.5	94.7	0.86	0.95
TB1PA3	0.78	12.2	97.9	0.89	0.98	0.78	12.2	97.9	0.89	0.98
TB1PA4	0.76	10.4	96.8	0.87	0.97	0.75	10.4	96.8	0.87	0.97
TB2PA1	0.77	18.1	98.0	0.78	0.99	0.77	16.7	97.8	0.85	0.98
TB2PA2	0.73	11.1	96.3	0.85	0.97	0.72	11.1	96.3	0.85	0.97
TB2PA3	0.75	17.6	98.5	0.73	0.99	0.74	19.0	98.6	0.69	0.99
TB2PA4	0.69	15.0	96.8	0.71	0.98	0.68	15.0	96.8	0.71	0.98

Table 5.7: Statistics for F1 and MCC

sensitivity and specificity. These can be compared to 'optimum' values provided by a ROC curve. Table 5.6 provides an example of this for TB1. These values, particularly sensitivity and specificity, can be compared with those actually achieved by the three thresholding methods in table 5.4.

According to Jeni et al. [62], if the data distributions are imbalanced, as is mainly the case here, most performance metrics are subject to an attenuation. They suggest the ROC curve to be an exception to this. However it was also found that while ROC is unaffected by imbalances it may also mask poor performance. This can be identified by the F1 score. In general if the sensitivity and specificity values resulting from a ROC analysis are accompanied by a high F1 score there can be confidence in the results. Similarly a poor ROC and poor F1 suggest a poor metric. If, however, a ROC analysis produces high sensitivity and specificity values, a correspondingly poor F1 does not necessarily mean that the ROC result is invalid. In this case some further consideration may be necessary. All results involving ROC analysis are accompanied with a corresponding F1 score. Although this may occasionally be used for comparisons itself it is, in most cases, used as an indication of the validity of the ROC analysis. For example, the F1 figures in table 5.6 are for the corresponding ROC threshold. In this instance the values indicate that for images of peak A-scans the corresponding ROC is satisfactory. For greater confidence in the use of peak A-scan values as a means of classification these should ideally be higher.

5.13 Fault location and sizing

Image segmentation provides a visual indication of the location of any anomaly. For each instance automatic location and sizing first requires extraction of each feature from the

background. In the case of black and white images one of the simplest approaches to this is through blob detection [64]. In essence this refers to the process of identifying regions of interconnected foreground pixels. For this application use is made of a Matlab in-built function (regionprops) which produces statistics defining each anomaly, including its size (number of pixels) and location (by means of a bounding rectangle).

Figure 5.8 illustrates two examples of this using different thresholding (KI and Otsu). At this stage, after blob detection, any area containing less than 5 pixels is omitted on the basis that is more likely to be noise. The Otsu threshold provides an indication of all anomalies. In each case the lower threshold of the KI technique gives a greater emphasis to each. In particular single anomalies are not as readily split, as is the case with the Otsu method. It is also indicated that the KI threshold provides more false positives. However some of these actually represent regions containing unintentional anomalies that may warrant further investigation.
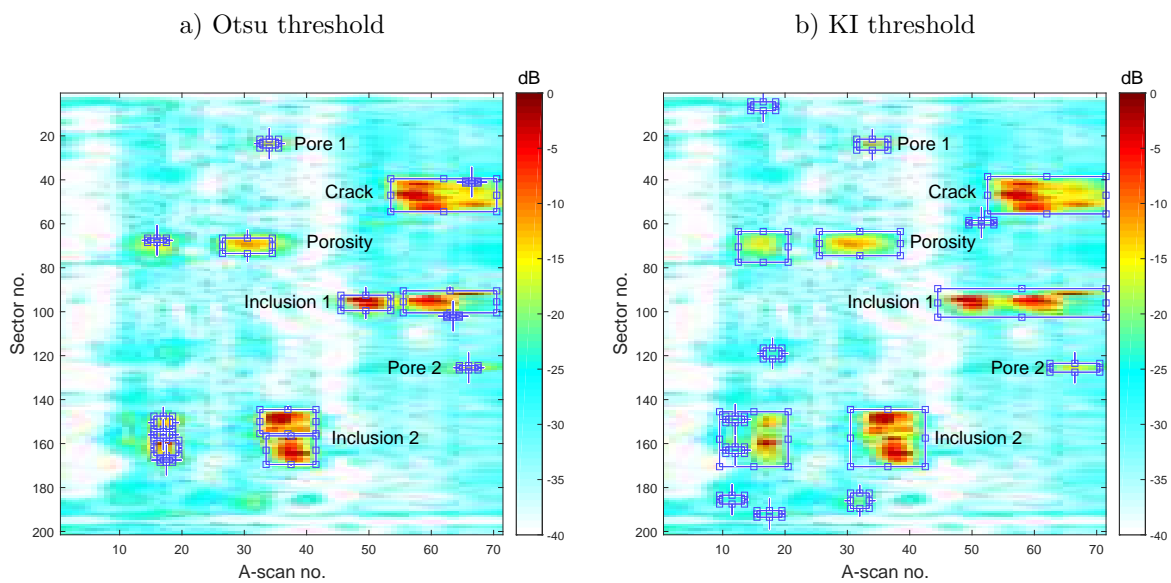
Figure 5.8: Fault locations for TB1PA

The primary aim of this work is to locate anomalies in large quantities of data. A future extension to this could be to classify each fault. Initially this might be to make a binary decision of an anomaly as being either critical or non-critical, the decision being based on some combination of the anomaly's size, location and shape. An extension of this might then be to enumerate the fault (e.g. crack, pore, porosity etc.).

Any characterisation will likely require more information on each anomaly, such as that

in figure 5.9. This provides illustrations of two surface profiles (magnitude and depth) for each anomaly in TB1PA1. Each profile is from the set of A-scans corresponding to each anomaly. The magnitude profile illustrates the peak A-scan value within the set, whilst the depth profile is the distance along the A-scan that the peak occurs. To aid comparisons the magnitude values are normalised with respect to the data set's global peak value. The depth units are in centimetres with the datum being the point at which the A-scan's gate turns on, figure 5.1.b. It is understood that, to date, no attempt has been made at characterisation using this type of data. To investigate such an approach requires far more examples than are available from the two test blocks used here.

5.14   Signal-to-noise ratio

The ability of any technique to detect an anomaly is, in part, dependent on the anomaly's prominence in comparison to its background. For example, consider the two instances of porosity, figure 5.8.a (sectors 65 to 75). The Otsu threshold is able to give a strong indication of the most prominent instance but captures only a small number of pixels corresponding to the weaker second instance. In fact any further post processing of this data might conclude, based on the pixel area, that this second instance is noise. More of this second instance is captured in figure 5.8.b. However, this is due to the lower KI threshold, the tradeoff being a greater number of false positives.

One approach to improving discrimination is to improve the Signal-to-Noise Ratio (SNR) of each anomaly. This chapter does not consider how this might be achieved. However this is an appropriate point to introduce the method of measurement and, for future reference, list the SNR of each anomaly in TB1PA1.

Various methods of measuring the SNR of an image feature are available [65]. This work adopts a common approach for use with TFM images (Villaverde et al. [66]). Its definition is:-

$$SNR_{dB} = 20 \log_{10} \left(\frac{I_{max}}{\sqrt{I^2_{noise}(\mathbf{r})}} \right) \tag{5.10}$$

where I_{max} is the maximum amplitude of the anomaly and $I^2_{noise}(\mathbf{r})$ is the mean square value of the background. The maximum amplitude is from a rectangular outline drawn around the anomaly. Similarly the RMS is from a set of intensities inside a rectangle drawn around an adjacent area of background.
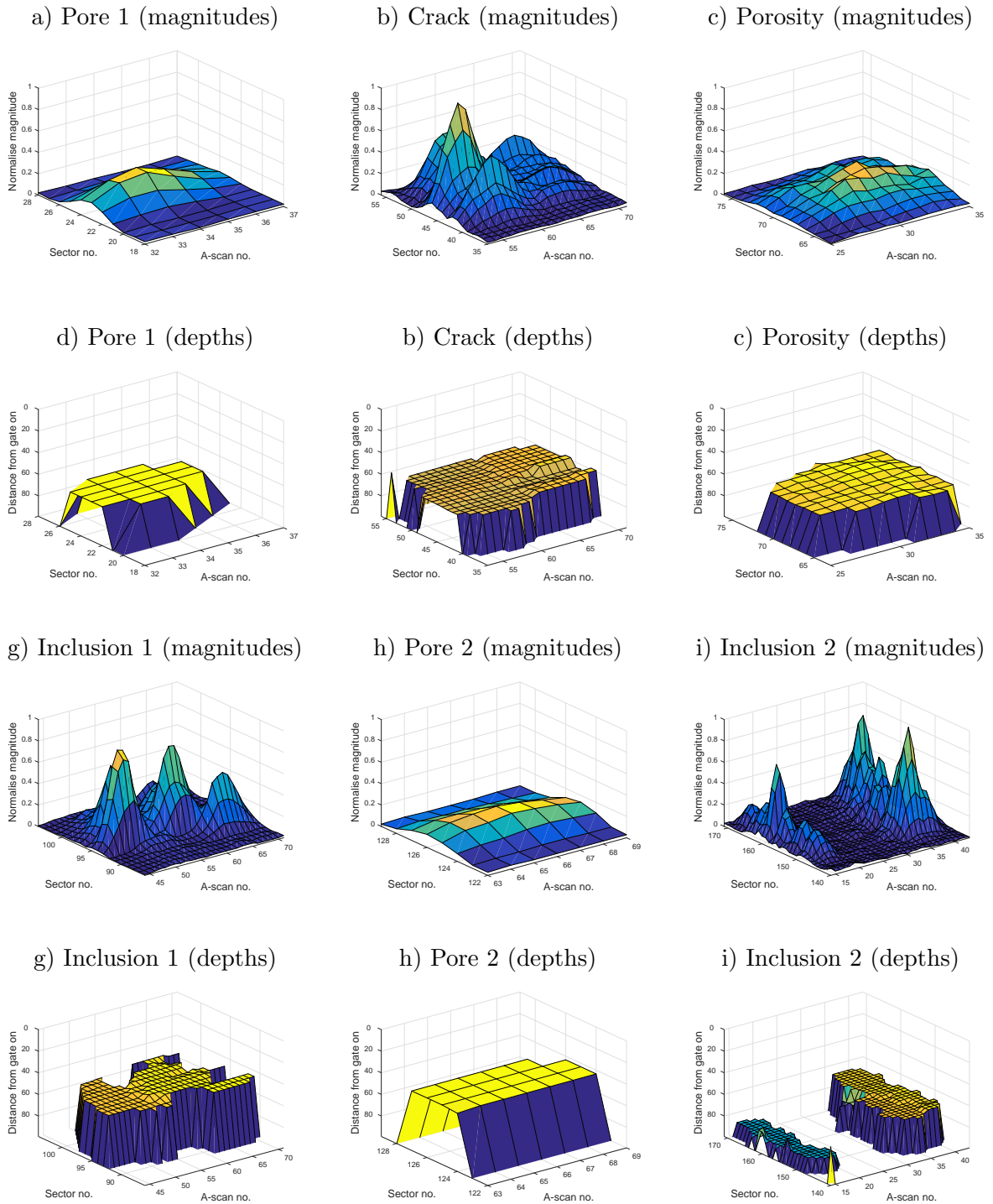
Figure 5.9: Magnitude and depth profiles of each anomaly from TB1PA1

For consistency this method will also be used for measuring the SNR of anomalies within the sectorial data sets. In this case the background selected is from the same A-scans as those containing the anomaly. This ensures that, with a shift along the weld, the two regions have the same viewpoint of the weld. For TB1PA1 figure 5.10 illustrates the regions, table 5.8 lists the corresponding SNRs. All future measurements will use the same areas.

a) All regions   b) Instance 1 of inclusion 2
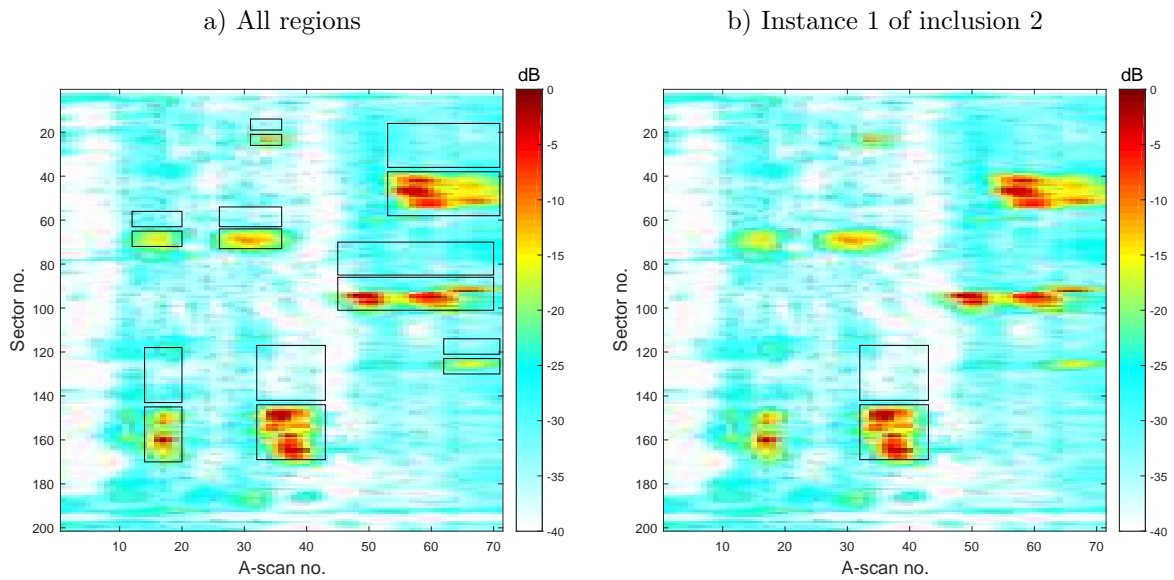
Figure 5.10: Regions of interest for SNR calculations

Fault Instance	SNR (dB)
Pore 1	23.67
Crack	31.25
Porosity I1	24.22
Inclusion 1	29.58
Pore 2	18.78
Inclusion 2 I1	35.82
Porosity I2	18.54
Inclusion2 I2	25.67

Table 5.8: SNR's using A-scan peaks

Figure 5.11 gives an example image from each TFM data set. In each case the pair of rectangles defines the region used to determine a features SNR. For possible future reference, table 5.9 lists the SNRs of each anomaly taken directly from its respective TFM image.

a) TFM1 b) TFM2 c) TFM3
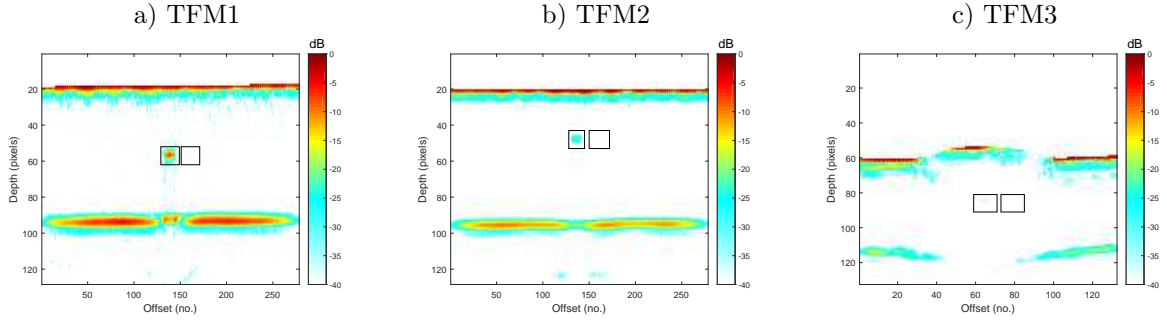
Figure 5.11: Example TFM images and areas for SNR calculations

Test Piece	SNR (dB) Fault no.				
	F1	F2	F3	F4	F5
TFM1	36.45	38.52	39.74	–	–
TFM2	28.69	31.77	35.51	–	–
TFM3	29.82	19.53	21.51	18.23	22.2

Table 5.9: SNR's using A-scan peaks

5.15 Immersion images

In the previous set up, using a wedged contact probe, the region of interest is some distance away from the point at which the ultrasonic beam crosses the front wall. The dominating front wall echo is now easily omitted from each A-scan by turning on the receiver after the return of these strong reflections. This is demonstrated in figure 5.1.b where the image starts some distance away from the probe. Only at this point do the recordings start (gate on). The end of each A-scan represents the point at which the gate turns off and recording stops.

In the case of these normally incident immersion tests, the front and back walls become dominating features. This is a direct consequence in the difference in acoustic impedance between water and the copper-nickel test piece. Under the assumption that the acoustic impedance of the test piece is similar to that of steel, figure 2.3 indicates that over 80% of the longitudinal wave's intensity is reflected from the front wall. Unlike the case of the oblique contact probe it is not possible to remove this large reflection without losing potentially critical surface information. In particular reflections from anomalies close to, or breaking through, the front or back walls are likely to be contained within these dominant reflections. In areas containing a weld cap or root, detection of such faults can be critical.

5.16 Thresholding immersion images

Accepting that thresholding is unlikely to detect surface anomalies reliably, consideration is given to the technique as a method of detecting anomalies within the material body. Following the sectorial data example, the first step is to obtain a two dimensional image. Previously this reduction to two dimensions was by taking the maximum value in each A-scan. With TFM images, conversion to two dimensions is possible from three viewpoints, figure 5.12.a.

In the absence of a front wall a plan view, figure 5.12.b, is the most similar orientation to that of the sectorial data. At the expense of depth information, this would have the potential to locate anomalies according to slice numbers (position along the weld) and offset (distance across the image). However in this immersion case it is clearly evident that surface reflections prohibit this.

Each pixel point in figures 5.12.c and d is from the maximum value in the respective directions. Once again there are significant problems with each of these. In particular it is apparent that, along all TFM images, there is some shift in vertical direction. This effectively thickens and blurs the front wall image, potentially obscuring anomalies. In the case of the side view the shallow anomaly (F4) is completely buried within the back wall. For the end view it is difficult to associate any anomaly with its respective TFM image. In practice there may be many hundreds of TFM images and one anomaly may entirely overlap a smaller anomaly at a deeper level. Although anomalies are evident in this image they are at a reduced SNR. Table 5.10 gives this comparison. The overall conclusion has to be that, in this case, reduction to a two dimensional image for thresholding is not a viable option.

Test Case	SNR (dB) Fault no.				
	F1	F2	F3	F4	F5
Individual Images	29.82	19.53	21.51	18.23	22.2
Maximum End View	22.73	17.25	13.55	8.14	15.45

Table 5.10: TFM3 SNRs from end view

5.17 Chapter summary

This chapter has investigated automatic anomaly detection using a 2D image (plan view) of the weld. In the case of sectorial data the value of each pixel is taken as the peak

a) TFM viewpoints
b) Example plan view (TFM3)
c) Example end view
d) Example side view (TFM3)
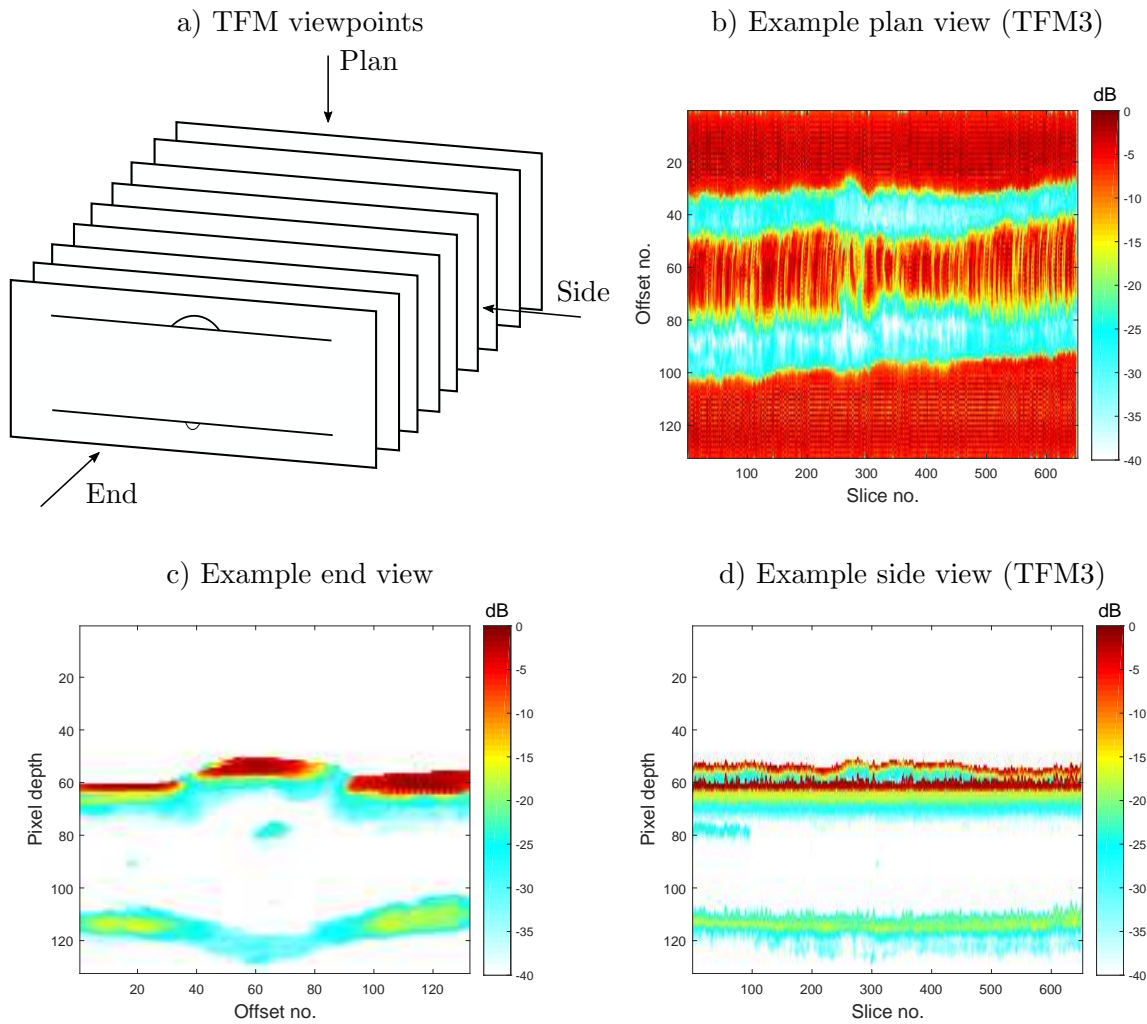
Figure 5.12: TFM view points

value from the corresponding A-scan. For the TFM images, which contain a front wall, the maximum value of each column vector (for example figure 5.11) produces only a pattern representative of the specimen's surface (figure 5.12). Therefore the presented results are for the sectorial data sets only. For these instances automatic image segmentation is by thresholding and three methods have been introduced. Assessment of each method is based on a comparison of the binary image after thresholding with those of the respective ground truth (figures 4.12 and 4.13).

After comparison with the ground truth, measures of the effectiveness of each threshold value are through the use of a confusion matrix (section 5.10). In this context one of the simplest measures is accuracy, for which MCE is closely related. It is demonstrated that

without a high value of accuracy classification is poor. However high accuracy alone does not necessarily indicate good classification (table 5.5). Alternative measures are based on specificity and sensitivity. For perfect classification both of these measures are 1 (ie. 100%). In practice there is a trade-off between the two. Typically a high threshold level has a tendency to reduce sensitivity whilst increasing specificity; in contrast a low threshold value has the opposite effect. To determine a theoretically 'optimum' trade-off, use is made of the ROC curve (section 5.11.1). To automate the tests, a software test bench which compares an image after thresholding with the corresponding ground truth has been developed. Throughout this work all test results will now use this test bench and ground truth data.

It has also been demonstrated that segmenting an image based on the maximum A-scan values has some potential for anomaly detection. However results (table 5.4) indicate relatively poor sensitivity for all thresholding methods. The ROC results (table 5.6) demonstrate that a more 'optimal' threshold value is possible although, in most cases, the results remain unsatisfactory.

The following chapter investigates a method of improving classification using Principal Component Analysis (PCA). This is a multi-dimensional technique which identifies correlations between variables, thereby reducing the original data set's dimensionality. This results not only in significant improvements to classification but, more notably, it is shown to significantly reduce the dominance of the front wall. This allows the TFM data (or any other data with a dominating front wall) to be analysed in a similar way to that of the sectorial data.

Further chapters will demonstrate methods of automatically improving the robustness of the PCA procedure. In all cases the thresholding methods outlined here are used, as are the methods of assessing and reporting the respective classification results.

Chapter 6

PCA trials

The previous chapter evaluated thresholding as a means of locating potential anomalies within a test block. This chapter furthers the anomaly detection objective through the use of a multivariate technique known as Principal Component Analysis (PCA) [67]. Known also as the Karhunen-Loéve transform, one attraction of PCA is its ability to reduce a high dimensional data set to an approximation with a smaller number of principal components.

One of the earliest reported attempts at analysing ultrasonic data using PCA is that reported by Bae et al. [21]. Their primary objective was to investigate PCA for fault classification. The work indicates some promise of the technique, but results were limited to the inspection of a short section of a single pipe weld. Further investigations on fault characterisation by Lingvall and Stepinski [68] suggest that some defects are hard to distinguish from individual signals alone and they further suggest that more advanced features, including PCA, be used.

This chapter provides an overview of PCA. It then illustrates methods of applying the technique to locating weld anomalies from ultrasonic phased array data. Although the primary aim is that of fault identification, indications of how the technique may also be used for characterisation are also presented. However, because faults will exhibit themselves in different manners, depending on the actual application, the primary aim of the work remains that of automatically identifying anomalous behaviour in a given geometrical configuration.

6.1 Framework for PCA

In general it is possible to express experimental data in a table where each row represents a single measurement (e.g. temperature, voltage, pressure) and each column a single recording of all measurements. For this scenario the number of rows represents the number of variables, or dimensions, with each column representing a single observation, or trial. One objective of PCA is to reduce the number of dimensions of the data for further exploration and analysis. The process is analogous to assigning each group of variables with good correlations to a particular component, the principal component being the one with highest variance. When performing PCA there is always the hope that the variances of most of the principal components are negligible. If this is the case it is then possible to adequately describe the variation by the small number of major principal components thus achieving some degree of economy.

6.2 PCA

More precisely let X denote the a $p \times n$ data matrix. Each column, n, represents an observation of a p dimensional vector:-

$$X = \begin{bmatrix} x_1 & x_2 & \dots & x_p \end{bmatrix}^T \equiv \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1,n} \\ x_{21} & x_{22} & \dots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p1} & x_{p2} & \dots & x_{p,n} \end{bmatrix} \tag{6.1}$$

PCA attempts to explain the variance-covariance structure of the p variables of X through a smaller number of new variables (principal components) that are linear combinations of the original variables. The principal components are therefore linear combinations of the p random variables $X_1, X_2, \dots, X_p$. They have three important properties:-

- There is no correlation between each principal component.

- The principal components are ordered in terms of descending variance.

- The total variation in the principal components is the same as that in the original data.

The principal components are obtained through analysis of the covariance matrix. If vector X has mean $\mu = (\mu_1, \mu_2, \dots, \mu_p)^T$, then the covariance matrix of X ($Cov(X)$) is the square

$p \times p$ symmetric matrix:-

$$\Sigma \equiv Cov(X) = \mathbb{E}[(X - \mu)(X - \mu)^T] \tag{6.2}$$

The covariances between the ij^{th} element $(i \neq j)$ of Σ is:-

$$\sigma_{ij} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] \equiv cov(X_i, X_j),$$

and the diagonal $(i = j)$ elements of Σ are the variance of the components of the population matrix X:-

$$\sigma_{ii} = \mathbb{E}[(X_i - \mu_i)^2] \equiv var(X_i)$$

and are consequently non-negative.

Since the covariance matrix is symmetric, $(\sigma_{ij} = \sigma_{ji})$, its eigenvalues are all real and positive; additionally the eigenvectors belonging to each eigenvalue are orthogonal. The eigenvector (e_i) with the largest eigenvalue (λ_i) is the first principal component of the data set. Ordering the eigenvectors by their eigenvalues (from highest to lowest) gives the principal components in order of significance.

The matrix of ordered orthonormal eigenvectors:-

$$E = [e_1, e_2, \ldots, e_p] \tag{6.3}$$

forms a basis for projecting the original data set into the new data space whilst the transformation:-

$$Y = E^T.X \tag{6.4}$$

rotates the original data onto a set of axes defined by the eigenvectors. For example, if z represents the zero centred observation of $x = (x_1, x_2, \ldots x_p)^T$, then the ith principal component of the projection is:-

$$y_i = e_i z = e_{i1} z_1 + e_{i2} z_2 + \cdots + e_{ip} z_p \quad i = 1, 2, \ldots, p \tag{6.5}$$

where $e_i = (e_{i1}, e_{i2}, \ldots, e_{ip})^T$ is the ith eigenvector. The kth zero centred observation is defined as:-

$$z_k = x_k - \bar{x}_k \tag{6.6}$$

Typically a projection will use only the first m principal components $(m < p)$.

The projected data, Y, is returned to the original co-ordinate space by reversing the rota-

tion of equation 6.4:-

$$Xrecovered = E.Y \tag{6.7}$$

Depending on the application it may now be necessary to reverse the effect of the zero centering, equation 6.6. This is easily achieved by adding the average of each variable, $\bar{x}_k$ to each column of $Xrecovered$.

6.3 Geometric description of PCA

As a simple illustration of PCA figure 6.1 illustrates a plot of a 2 dimensional (x_1, x_2) data set (with observations $n = 50$). In this case the two variables are each zero-centred (equation 6.6) before arranging into a 2×50 matrix (X in equation 6.1). The covariance matrix (equation 6.2) defines both the spread (variance) and the orientation (covariance) of the data. The largest eigenvector of the covariance matrix points in the direction of the largest variance, its magnitude being given by its associated eigenvalue. The second largest eigenvector is always orthogonal to the first and gives the direction of the second largest variance. Figure 6.1.a illustrates the two eigenvectors (e_1, e_2) associated with this two dimensional data set. The length of the bold line superimposed on each eigenvector represents the respective eigenvalue. In the following it is assumed that the eigenvectors are ordered with the first eigenvector, e_1 having the largest eigenvalue.

In figure 6.1.b the original data, x_1, x_2, is zero centred (equation 6.6 to produce z_1, z_2. The projected data points y_1, y_2 (equation 6.5) are with respect to the axis defined by the two eigenvectors, e_1, e_2. No information is lost and there is no change. If however, the projection uses only the first eigenvector, so that $y_2 = 0$, then the projection y_1 is along e_1 only; all the points lie on e_1. Information representing the variance with respect to the axis e_2 is lost.

6.4 Inverse sample covariance matrix

An alternative geometric description of principal component analysis is given by Gnanadesikan [69]. In this case the inverse of the covariance matrix is employed as the matrix of quadratic form, which defines a family of concentric ellipsoids centred on the sample centre of gravity. That is, the equations:-

$$(X - \mu)\Sigma^{-1}(X - \mu)^T = c \tag{6.8}$$

a) Eigenvectors e_1, e_2 b) Projection onto e_1
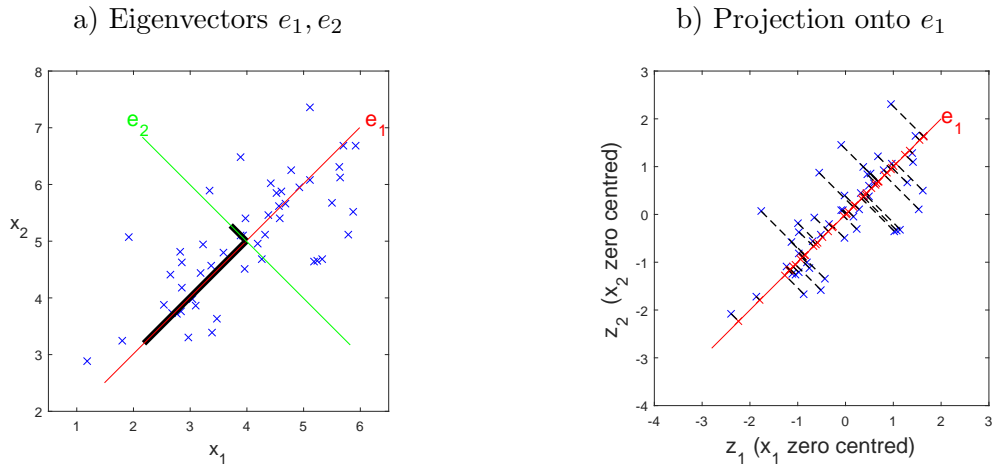
Figure 6.1: Geometric description of principal components (PCs)

a) Original sector scan   b) Recovery from 3 PCs

c) Recovery from 10 PCs   d) Recovery from 20 PCs
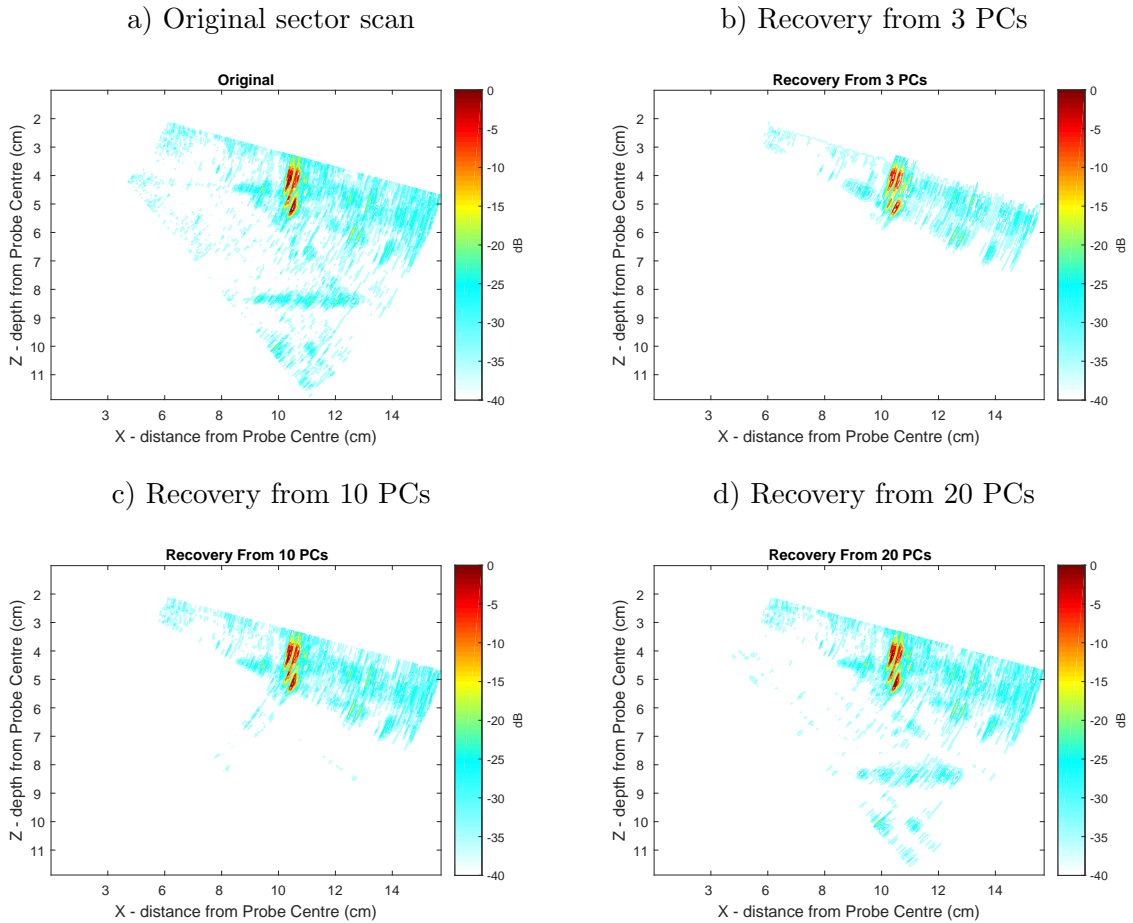
Figure 6.2: Data compression and de-noising using PCA

for a range of non negative values of c, define a family of concentric ellipses in the p-dimensional space of X. The principal component transformation of the data is the projection of observations onto the principal axes of this family. For example in the two dimensional case the original co-ordinates (X_1, X_2) are transformed by a shift of origin to the sample means $(\bar{X}_1, \bar{X}_2)$ followed by a rigid rotation that yields the principal component co-ordinates z_1 and z_2.

6.5 PCA for data compression and de-noising

With PCA the assumption is that the variance of most of the variables will be so low as to be negligible. If this is the case most of the data variation is adequately described by a small number of principal components. For example, consider the sector scan illustrated in figure 6.2a. This data set contains 71 A-scans each with 1544 samples. Arranging this as a matrix X 71 mean centred variables and 1544 observations gives a covariance matrix of size 71×71. A singular value decomposition producing of this gives a maximum of 71 principal components or eigenvectors.

If E is corresponding $71 \times P$ matrix containing the top P (magnitude ordered) eigenvectors, then the projection:-

$$Y = E^T X \tag{6.9}$$

reduces X to a matrix, Y, of size $P \times 1544$. Reversing the process:-

$$X_{rec} = E.Y \tag{6.10}$$

recovers the image but using only the chosen principal components. Figures 6.2.b-c, illustrate the projection and respective recovery of the original image using the top 3, 10 and 20 principal components respectively.

Data compression is achieved by reducing the original 71×1544 image to two matrices, Y of size $P \times 1544$ and E of size $71 \times P$. Similarly the removal of principal components containing less correlated variations de-noises the image. In some applications these principal components may contain important information. In cases such as this example they are more likely to be representative of noise.

6.6 PCA projections

The previous example demonstrates the use of PCA as a method of data compression and de-noising. The projections of the original data set into the lower dimensional representation (equations 6.5 and 6.9) are of particular interest to this work. The purpose of this section is to demonstrate how PCA can be used for fault detection. This starts with an example that uses a simple and well known reference data set (the Fisher's or Anderson's data set [70]). Collected by Anderson the data lists 4 measurements (sepal length, sepal width, petal length and petal width) taken from 3 species of Iris (setos, versicolor and virginica). In total there are 150 samples with 50 examples of each species. The objective is to find some feature of the data set that will separate the species. Then, given a set of measurements from a new sample, it should be possible to identify the species to which it belongs with some degree of confidence. One simple way of separating the species is to create scatter plots of any two variables to see if any pattern exists. From the four measurements six scatter plots are possible with some giving better discrimination than others. As an alternative PCA provides a structured and straightforward approach to creating simpler plots that capture most of the information directly. For example, figure 6.3 illustrates the result of projecting a data set against the first two and first three principal components.

It is known a-priori which species a particular set of measurements, or observation, belongs to, so it is possible to assign a colour to each projection on the plots. Figures 6.3.a (which captures about 85% of the variance) and 6.3.b (which captures about 95% of the variance) illustrate that the three species form separate clusters in the lower dimensional space. The projections can now be used for pattern recognition. For example, if a set of measurements from a new, unknown, species is projected against the existing set of eigenvectors the result is expected to be a point close to, or within, one of the existing clusters. This, in turn, identifies the species to which the unknown sample belongs. In some instances the new point may yield an ambiguous result or, in a more extreme case, if the point is far removed from the existing clusters, it is likely to be representative of a species not included in the original data set.

6.7 Application to ultrasonic inspection

Although the above example uses a straightforward data set it illustrates many of the features of the technique. For this work the objective of investigating PCA is to evaluate

a) Projection against first 2 PCs   b) Projection against first 3 PCs
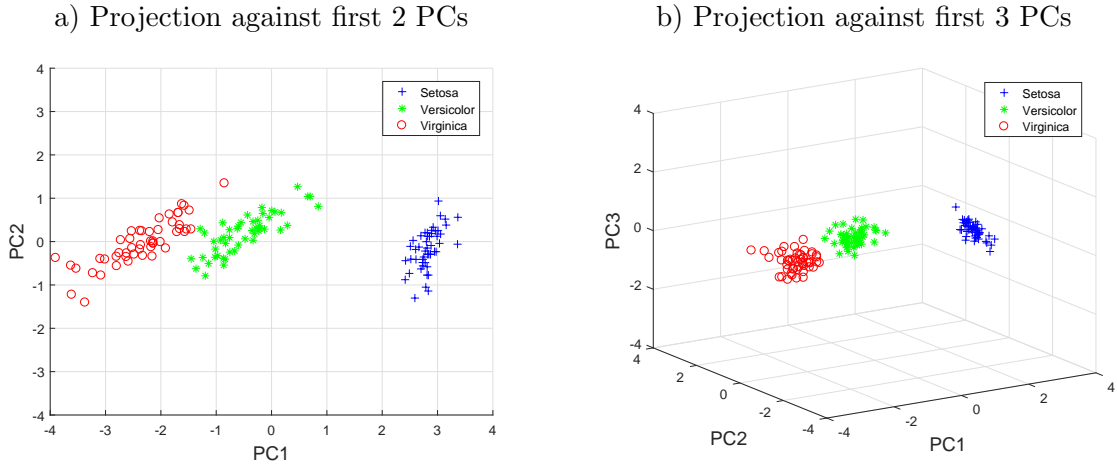
Figure 6.3: Example projections (from Fisher's iris data set)

its ability to classify a given ultrasonic test as either containing an anomaly or as being anomaly free. In the case of anomaly free inspections each data set will contain information representing background details only. All such data sets will be similar and their projections, in a reduced dimensional space, will belong to a single cluster. A collection of these anomaly free observations will be termed the training set. If the projection of a single unknown observation against the training set results in a point within the training set's cluster of observations, then the unknown observation is likely to be anomaly free. Alternatively if the projection is far removed from the cluster's centre, it indicates that the observation is likely to contain an anomaly.

Data from two types of ultrasonic test is under consideration, namely sector scans and TFM images. Although these are to be considered separately there is some commonality in each approach. Firstly, the training set of anomaly free observations needs to be organised in such a way that the projection of a single test observation is sensitive to any anomaly. Related to this will be the number of principal components to use. Methods of measuring the multidimensional distance of a test projection from either the centre of, or individual points of, the training cluster are needed, as are methods of evaluating any threshold distance.

6.8 Sector scan investigations

The work starts by considering sector scan data. The base data sets are those used previously from TB1 and TB2. Each test block has been scanned four times in directions

98

PA1 to PA4. There is a total of 1816 sector scans but only 1633 are valid. Of these about 70% (around 1100 sectors) are anomaly free. From a visual inspection of the fault maps (e.g. figure 6.8.a) 700 fault free sectors were isolated. Initially, training sets are randomly selected from this group. In the following the term 'manually selected training set' refers to observations taken from this set.

6.9 Training set organisation

Unlike the previous examples (section 6.6) where each observation is a single vector, an observation is now a sector image containing a matrix of A-scans and samples. These arrangements are common in many image recognition systems. An area where PCA has had noteable success is that of face recognition [71], [72]. One of the first presentations of the so called 'Eigenfaces' method was reported by Turk and Pentland [73]. Here each 2D image is converted to a single vector by concatenating each row such that, for example, an $n \times m$ matrix is converted to a single vector of length $n.m$. Each vector now makes a single observation.

6.9.1 Full sector training sets

For a sector scan consisting of 71 A-scans, each with 1544 samples, the overall column vector length, and number of dimensions (p), is 109624. The resulting covariance matrix, (6.8), is now 109624×109624. Although determining the eigenvectors of such a large matrix is not intractable on a modern computer it consumes considerable processing power and time. However in these cases the maximum number of training set observations, (n), limits the number of eigenvalues to $n-1$, all others having a value of zero. In such cases a computationally easier result proposed by Turk and Pentland [74] is to solve the transpose of the data matrix (leading to a $n \times n$ covariance matrix). This produces the n eigenvectors of the training set.

6.9.2 Constant A-scan training sets

An alternative to converting each sector into a single vector observation is to consider a group 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 6.4 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.
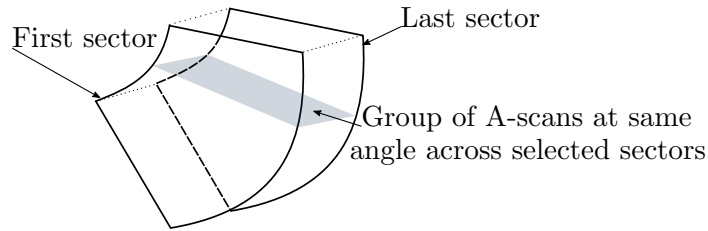
Figure 6.4: Training set as A-scans at constant angle

In the following both approaches are attempted. Results are presented initially in terms of the technique's ability to accurately detect anomalies in the test blocks and the virtual test pieces. The method is then extended to TFM data sets.

6.10 Investigative studies

To begin, some consideration needs to be given to the following:-

- Number of PCs to use: the principal components are ordered in terms of their variance, or eigenvalue. As indicated by figure 6.2, including more principal components increases the accuracy of the projection. Retaining too few PCs may cause features to be missed; retaining too many may result in false calls and produce an overly complex system. The decision on how many principal components to include is once again a matter of some judgement and needs some investigation.

- Training set selection: this has to be representative of the fault free background. An estimate of the background improves as the number of examples increases. However increasing the training set size contributes to the problem of automatically selecting representative, anomaly free, observations. In practice these are not known a-priori. Inadvertent selection of observations with anomalies detracts from the training set's accuracy. In the case of sector scans from the two test blocks, 700 fault free observations were previously isolated (section 6.8). These now facilitate an early evaluation of the technique. More specifically they will provide an indication of the performance, in terms of accuracy and efficiency, as well as answering more objective questions such as the minimum number of observations to include. The problem of training set

creation from new data is considered later.

6.11   Scree plots

Section 6.5 demonstrates that when reconstructing an image from reduced dimensional space the larger the number of PCs used the smaller is the reconstruction error. In figure 6.2 the images contain an anomaly; this is responsible for much of the variance. As a consequence recovery from just 3 PCs provides much information of the anomaly and very little of the remaining background variance. To reconstruct more of the background variance requires more PCs. This suggests that a small number of the first few PCs are sufficient to discriminate between a background only image and an image containing an anomaly.

A measure of the variance explained by each PC is given by the respective eigenvalue. A plot of eigenvalues for each PC is known as a scree plot. Two examples of scree plots for training sets of anomaly free sectors are given in figure 6.5.a and b, the main difference between the two being the number of observations (i.e. $n = 50$ and 700 for a and b respectively). In both cases the data is from a random selection of the manually selected anomaly free sectors. For comparison the plots have the same scale and are limited to the first 49 principal components. Both plots are similar, the main difference being that for 50 observations all of the variation is accounted for whilst in the case of 700 observations some variation (not included) is contained within the principal components of 50 and above. However these are diminishing from an already low value; they are subsequently ignored and otherwise the two curves are similar.

Figures 6.5.c and d represent scree plots for training sets consisting of fixed angle A-scans. Scree plots frequently contain two characteristic asymptotes meeting at an elbow. The broken stick model [75] uses this elbow, or point of inflection, to suggest a limit on the number of PCs to use. In the four instances here the elbow is not always distinct. However it is suggested that as few as 5 PCs might be sufficient for either type of training set. This is particularly the case where the larger number of observations is used. In this case the scree plot for each type of training set does not offer any suggestion as to which (full sector or constant A-scan) might be the better.
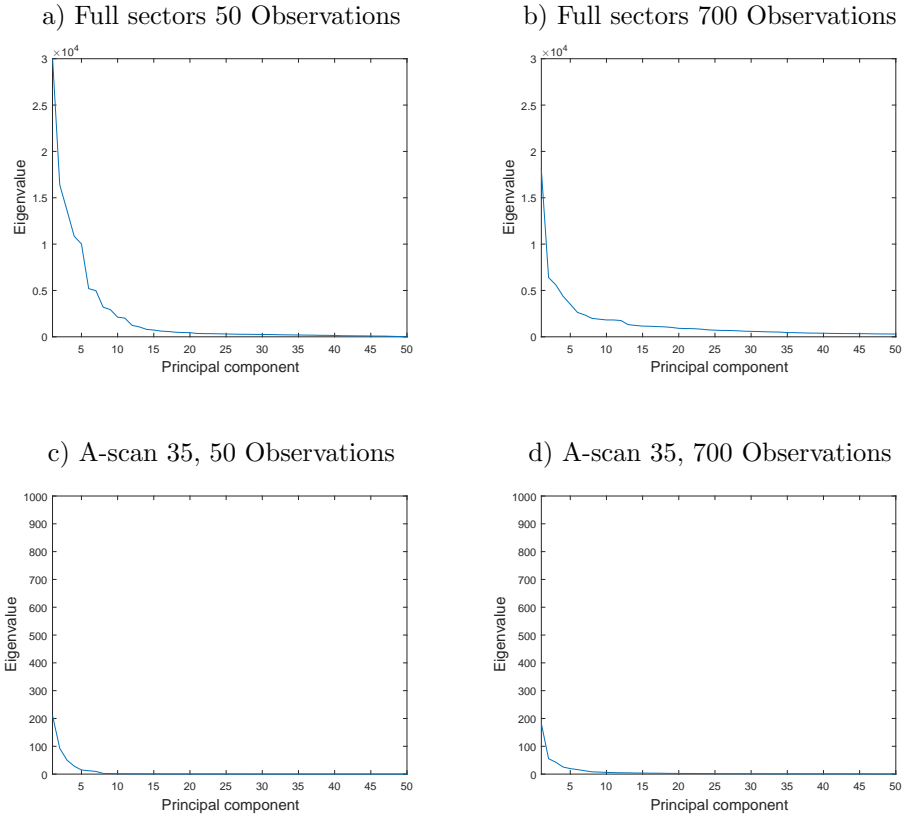
Figure 6.5: Scree charts for anomaly free training sets

6.12 High dimensional low sample size data

Most text book examples of PCA use data sets where the number of observations (n) is greater than the number of variables (p). Typical of these is the Fisher Iris data set (section 6.6) which has dimensionality $p = 4$ with $n = 150$ observations. In such cases using all principal components permits full reconstruction of the original data set. By contrast the ultrasonic systems under consideration here contain many more dimensions than observations, that is $p >> n$. For these High Dimensional Low Sample Size (HDLSS) systems the maximum number of principal components is limited to $n - 1$. This limit suggests that some reconstruction error is now inevitable, even using all available PCs, potentially compromising the PCA approach. A further limitation is that the upper limit of $n - 1$ principal components is not always realisable, in particular as $p_i \rightarrow (n - 1)$, the corresponding eigenvalue (λ) reduces such that, $\lambda_i \rightarrow 0$. Very low values of λ are not useful and in some cases their numeric values are outside the precision of the computer. In practice then the number of useable PCs may be much less than p and less than $n-1$. Despite these limitations evidence of the successful use of PCA in HDSLL applications is available [76] and [77]. More specifically a study of the HDSLL case by Jung and Marron [78] suggests that if the first few eigenvalues are large, compared with those remaining, the corresponding estimated eigenvectors converge to the appropriate subspace. All of the previous scree charts comply to this constraint suggesting that the technique remains valid. Further evidence of this is provided by figure 6.2. Here figure 6.2.b illustrates that the anomaly is almost completely recovered within the first 3 PCs, whilst a significant proportion of the remaining sector is recovered using 20 PCs, 6.2.d. In this work, whenever $p > (n - 1)$ the PCA algorithm returns the maximum number of non-zero eigenvalues.

6.13 Full sector versus constant A-scan observations

In the case of ultrasonic sector scans two possible arrangements of observations have been suggested. As the example data contains only 700 background examples, both arrangements result in HDLSS data sets. This is also likely to be the case in practice where the likely scenario is for a full weld not to exceed $3m.$, with a typical sector scan spacing of $2mm$. Given this outcome some further comparison of the two arrangements is required.

The potential advantage of a full sector scan observation is that the entire sector is inspected by a single projection. The disadvantage is the large number of dimensions. In the case of constant angle A-scan observations the dimensions are immediately reduced to the number

of samples in the A-scan. However to fully test a sector the projection must be repeated for each A-scan angle using a specific set of eigenvectors. However, more important than this cost comparison is the ability of each approach to detect anomalies accurately.

At an elementary level anomaly detection using PCA relies on the difference in eigenvalue distributions between a set of anomaly free observations and a set containing an anomaly. In particular the distinctiveness of the anomaly free eigenvalue population from that of one containing an anomaly may give an indication of the sensitivity of the approach. An illustration of this is given by figures 6.6.a and b for full sector observations and figures 6.6.c and d for single angle A-scan observations. In addition to the scree chart each plot includes the cumulative variation explained by the increasing number of principal components.

Figures 6.6.a and b indicate that, in the case of an anomaly free sector, the first 5 PCs account for around 30% of the total variation, whilst in the presence of an anomaly the same number of PCs explain approximately 80% of the total variation. In the case of single A-scan projections, figures 6.6.c and d, indicate a similar discrepancy between the cumulative explanation of anomalous and anomaly free observations, except that here the difference is more pronounced, suggesting that the A-scan only approach is more sensitive than full sector projections. In practice, and in both cases, the difference between anomalous and anomaly free conditions will depend on the size of the anomaly.

6.14 Anomaly recognition

The previous discussion suggests that with an appropriate feature around 5 PCs might be sufficient to discriminate between an anomalous and anomaly free observation. For this illustrative discussion and to aid visualisation, the number of PCs is restricted to 2. To re-iterate, the basis of the technique is to select a number of anomaly free observations. These are then projected into a reduced dimensional space defined by the dominant eigenvectors of the training set's covariance matrix. The similarity of each background observation will result in the projection of a scatter plot containing a dominant cluster that is statistically representative of anomaly free observations. Using the same eigenvectors, the projection of a new, anomaly free test observation will result in a point that is not distinct from the training set's cluster. Alternatively, projection of an observation containing an anomaly will be to a point away from the training set's original cluster. This principal is demonstrated by figure 6.7.a which contains a scatter plot of 200 background sectors plotted against the first two principal components. Two test sectors are now projected against the same eigenvectors. The sector without anomaly lies within the main cluster whilst the

a) Full Sectors anomaly free

b) Full Sectors with anomaly

c) A-scan 35, anomaly free

d) A-scan 35, with anomaly
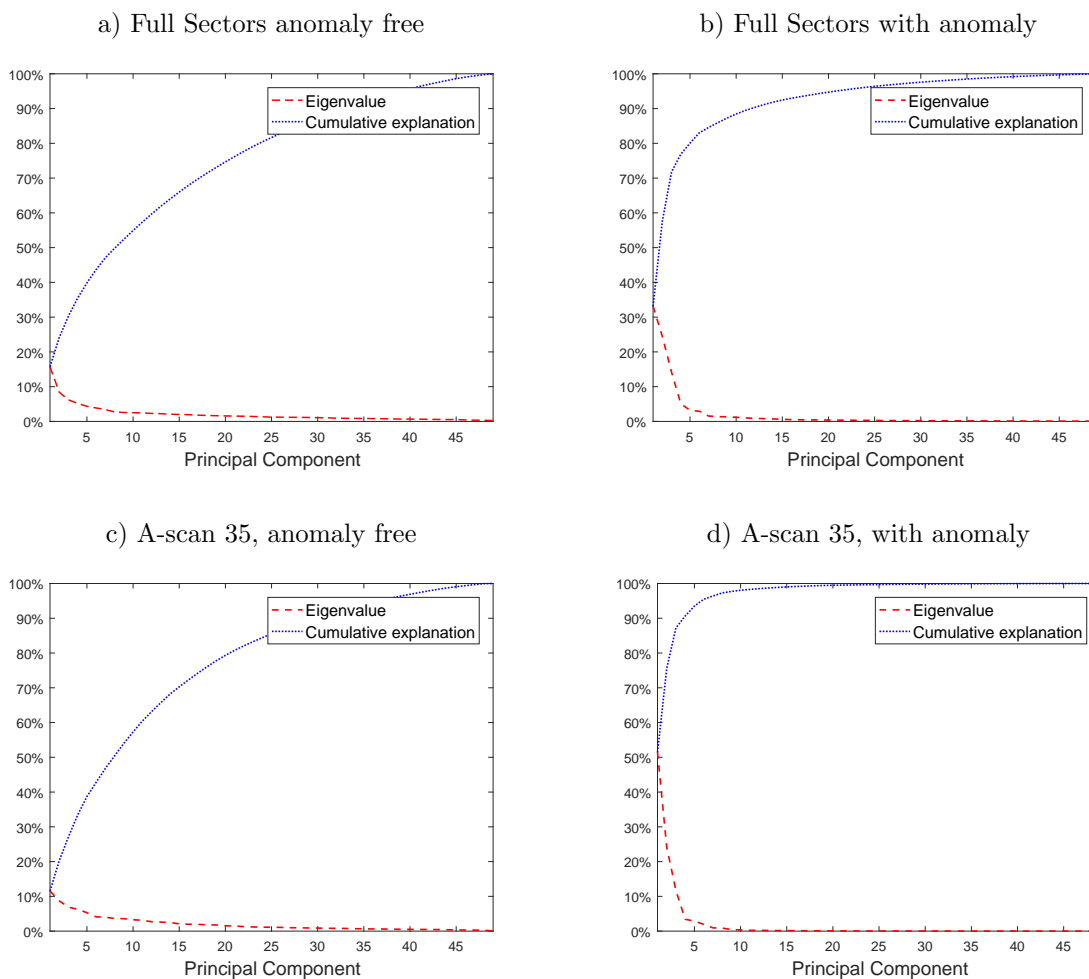
Figure 6.6: Example scree charts for 50 observations

sector containing an anomaly is distinctly outside the cluster.

6.15 Outlier detection

The literature refers to anomalous observations as outliers and the subject has been extensively studied [79] with many detection methods being proposed. These broadly fall into three categories, namely models based on statistic properties, distance and density.

Statistical models [80] and [81] are based on the assumption that the data has a distributional model and that knowledge of the distribution is available. The techniques are best suited to the univariate case, from which they originate, and are not so well suited to high

dimensional applications. If, however, the variables can be reduced to a small number of significant dimensions (for example through PCA), it may be possible to test that each projection conforms to its expected distributional model.

The problem of obtaining a distributional model remains, as does the fact that excluded, less significant, principal components may contain important information [82].

Distance models [83], [84] determine the distances between the observation of interest and its neighbours. If the result is above some predetermined threshold the observation is considered an outlier. No prior knowledge of the data distribution is required but a criticism of the technique [85] is that in the case of multi-clustered structures it may not be possible to select neighbours from the correct class, or cluster.

Density based methods [86] provide a method of overcoming the problems associated with multiple clusters. The basic concept is to determine the observation's degree of being an outlier, the Local Outlier Factor (LOF). It is local in that the degree depends on how isolated the object is with respect to any surrounding neighbourhoods or clusters.

For this application the consideration is that anomaly free data from a weld with constant geometry will project to a single cluster, or neighbourhood, in lower dimensional space. This removes the aforementioned limitation of the distance based approach. Similarly the advantage of density based methods does not apply.

6.16   Distance measures

For the multi-dimensional case it is necessary to detect whether or not a test point is within the 'cloud' of training set projections.

Figure 6.7.a illustrates the projection of 200 sectorial observations against their first two PCs; these are random selections from the 700 anomaly free observations. Two test sectors are now projected, using the training set's eigenvalues. One of these projects to a point close to the cluster's centre, giving confidence that it is anomaly free. The second projects to a point some distance away suggesting it is anomalous. The 95% confidence ellipse receives some attention in the following discussion which considers measures of the test observation's distance and its subsequent classification.

Of some interest from the figure is that the spread of PC1 is different from that of PC2. Assuming a normal distribution then for the univariate case it is common to determine the mean and standard deviation of the population. If a new observation is less than one

standard deviation from the mean, there is greater confidence that it originates from the same population than there is of a second observation that is, say, 3 standard deviations away.

For the multivariate case it is necessary to consider each dimension. Where the projections are zero mean centred the Euclidean distance of the anomaly from the centroid is, perhaps, the simplest measure. However the Euclidean distance does not take into account the variance of the projections along each principal component. For example, consider the two test observations in figure 6.7.b. Here, O_1 has a location $PC1 = 70, PC2 = 0$ (Euclidean distance $= 70$), and O_2 has location $PC1 = 0, PC2 = 60$ (Euclidean distance $= 60$). The Euclidean measure suggests that O_2 is the least likely to contain an anomaly. However this does not take into account the variance of each PC. By definition PC1 has greater variance than PC2. To account for this difference, in the two dimensional case, it is possible to draw an ellipse around the anomaly free training set with the ratio between major and minor axes determined by the respective spread of each PC. At this juncture the key note is that, despite having the largest Euclidean distance, the point O_1 falls within the boundary of the ellipse suggesting that, of the two, it is the more likely to be anomaly free.

6.17   Confidence ellipse

For bivariate data and given that $PC1$ and $PC2$ are uncorrelated the confidence ellipse makes the assumption that the data has a Gaussian distribution. The general equation of an ellipse with a major axis of length $2a$ and a minor axis of $2b$ is:-

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1 \tag{6.11}$$

In this case the length of the axes are defined by the standard deviations of PC1 and PC2. If the right hand side of the equation, 1, is replaced by S, a scale factor, then the size of the ellipse can be made to represent a chosen confidence level. Under the assumption that the scatter plot is from Gaussian data with zero covariance, then both the x and y values are also normally distributed, and the left hand side of the equation actually represents a sum of squares of independently normally distributed samples, meaning they follow a Chi-Square (χ^2) distribution [87]. In this case there are two unknowns or two degrees of freedom, and for a 95% confidence level the table of cumulative χ^2-distributions shows that the scale factor has a value of 5.991. The confidence ellipse for figure 6.7.a and b is therefore drawn with a major axis of $2\sigma_{PC1}\sqrt{5.991}$ and minor axis of $2\sigma_{PC2}\sqrt{5.991}$.

a) Distance example

b) Example bi-variate confidence ellipse
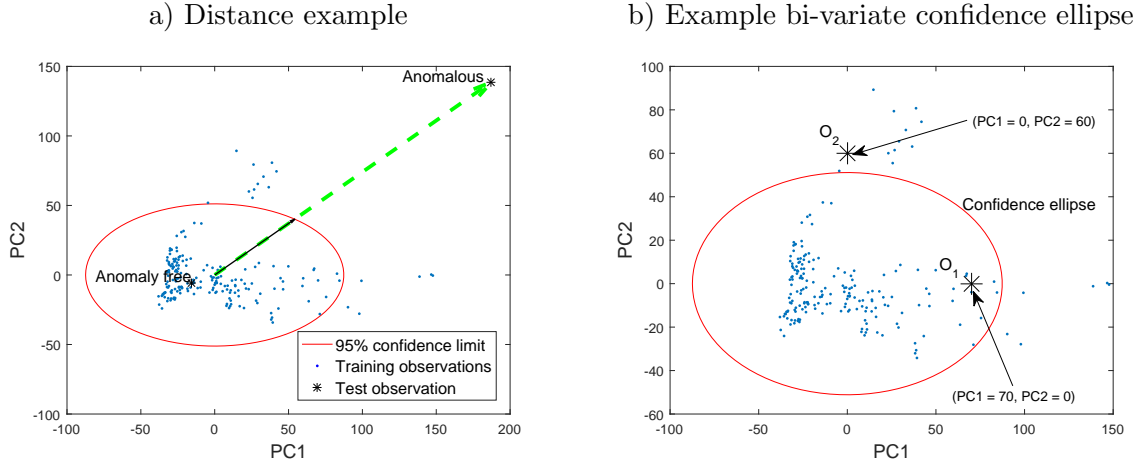
Figure 6.7: Examples of anomaly detection with PCA

The expectation is now that 95% of the training set observations will fall within the confidence ellipse. Similarly there is a 95% probability that any test observation falling inside the ellipse is free from anomaly. The further away the projection of a test observation is from the closest boundary point of the ellipse, the greater the probability that it contains an anomaly. To determine this point let A_{PC1} and A_{PC2} represent the location of the anomalous observation on the bivariate plot. The Euclidean distance (green dashed line) of the anomaly from the origin is given by $A = \sqrt{A_{PC1}^2 + A_{PC1}^2}$. This line intercepts the confidence ellipse at a point (black solid line on plot) given by:-

$$r = \frac{ab}{\sqrt{a^2 \sin^2 \theta + b^2 \cos^2 \theta}} \qquad (6.12)$$

where θ represents the angle between the horizontal axis and the anomaly, and a and b are the horizontal semi-axis and vertical semi-axis, respectively.

6.18 The Mahalanobis distance

The previous bivariate example can be extended to the multivariate case. In this sense the prediction ellipses become a multivariate generalisation of the univariate units of standard deviation. The probability contours can be used to define the so called Mahalanobis distance [88] which is a generalisation of the distance concept to p-dimensional correlated data. The method uses the covariance matrix to take into account the differences in variances and correlations between variables. Where the Mahalanobis is used to measure the

distance between a single observation and the centre of the population the distance is expressed as [89]:-

$$D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu) \tag{6.13}$$

where x is a p dimensional observation, μ is the population mean and Σ is the population covariance matrix. As previously suggested if the population is multivariate normally distributed then the values of D^2 follow a chi-squared distribution with p degrees of freedom. Once again this can be used as a threshold to determine if any observation is anomalous.

Jobson [90] shows that if y_i represents the standardised i^{th} principal component of an observation and λ_i the respective eigenvalue then the sum of squares:-

$$\sum_{i=1}^{p} \frac{y_i^2}{\lambda_i} = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \cdots + \frac{y_p^2}{\lambda_p} \tag{6.14}$$

is also equivalent to the Mahalanobis distance from the mean of the sample distribution and since the principal components are uncorrelated, for a large sample size, it follows that:-

$$\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i} = \frac{y_1^2}{\lambda_1} + \frac{y_2^2}{\lambda_2} + \cdots + \frac{y_q^2}{\lambda_g}, \quad q \leq p \tag{6.15}$$

has a χ^2 distribution [91] [89] with degree of freedom q. For this to be true all eigenvalues must be distinct, positive and ordered (i.e. $\lambda_1 > \lambda_2 > \cdots > \lambda_q > 0$).

6.19   Mahalanobis examples

Using the broken stick model, the scree diagrams, figures 6.6, suggest that most of the variation in a set of anomaly free observations (constant A-scan or full sector) is explained in the first 5 principal components. Figure 6.5, which compares anomaly free and anomalous observations, gives further confirmation of this. The following examples use the same 50, anomaly free observations, as training data. Test data is taken from TB1PA1 for which a reference map is given, figure 6.8.a. The map also illustrates two A-scans (35 and 66) which are used in the constant A-scan examples. Figure 6.8.b gives an indication of the sectors containing anomalies.

Figure 6.8.c describes the variation in the Mahalanobis (D^2) of A-scan 35, from each test sector, after its projection using 5 PCs. Prior to this the A-scan is mean centred in the usual way. To facilitate comparisons with the reference image each test result is rotated $90°$ so that the sector axis is vertical; on all test results the vertical red line indicates the

χ^2 95% confidence limit. Figure 6.8.d is a similar result for A-scan 66.

The result for A-scan 35 shows three clear peaks, each of which corresponds to the respective anomalous sectors crossed by A-scan 35. There is only a single distinct peak corresponding to the inclusion (sectors 90 to 100) of A-scan 66. The crack (sectors 40 to 50) and pore (sectors 120 to 130) have no significance. Although these results, particularly for A-scan 35, give some indications of faults, they are not good.

The example using full sector scans, figure 6.8.e is less convincing. Here D^2 for all sectors is weak. In the case of anomalous sectors the projections are mostly none distinct and are well below the confidence limit. A significant difference between the single A-scan and full sector approaches is the ratio of variables to observations. For the current images the length of an A-scan is 1544 samples. By comparison the length of a full sector, containing 71 A-scans, is 109,624 samples. Although the scree plots suggest 5 PCs a closer consideration of the accumulative explanation (figure 6.6.b) indicates that at most 80% of the cumulative variation is accounted for. This is in contrast to the fixed A-scan curve (figure 6.6.d) where the cumulative explanation is over 90% for the same number of PCs. For full sectors the corresponding curve suggests that at least 15 PCs are needed to approach 90% cumulative explanation. Although the results remain lacking, figure 6.8.f, indicates some improvement.

6.20   Limitations

In the case of HDLSS data sets the total number of PCs is limited by the number of observations $(n-1)$. If they are a true representation of background then increasing their number gives a more reliable representation of the anomaly free condition. Consequently increasing the number of observations is likely to lead to a better estimate of the number of PCs to use albeit at an increase in number. Increasing the observations also produces an increase in computational time to evaluate the eigenvectors. However this only needs to be done once. During detection, projection of a test observation is less demanding. A potentially more significant problem for a larger training set relates to its selection, particularly where the pool of potential candidates is limited. This subject is to be dealt with in a future chapter of this thesis. At present 700 manually selected fault free observations are available. This full set allows the use of up to 699 PCs. Selection of an appropriate number to use, requires some investigation. In particular, the previous 'broken stick' model is subjective and not amenable to automation. One alternative is the Kaiser stopping rule.
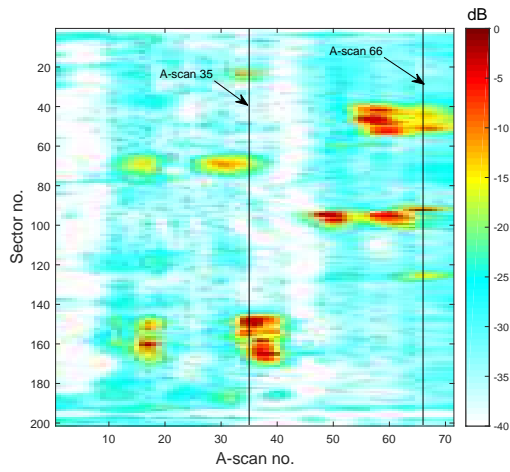
a) Reference image   b) Indicative anomaly locations
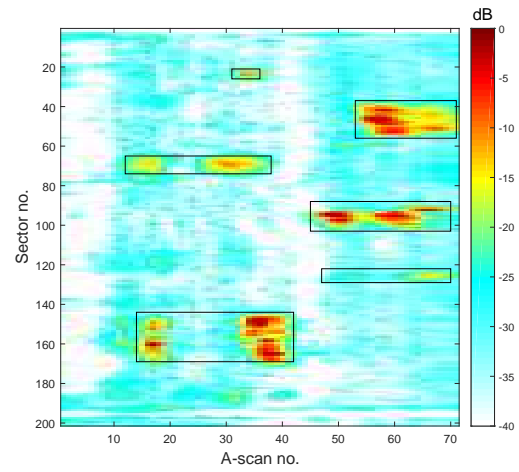
c) A-scan (35) only projections (5 PCs)   d) A-scan (66) only projections (5 PCs)
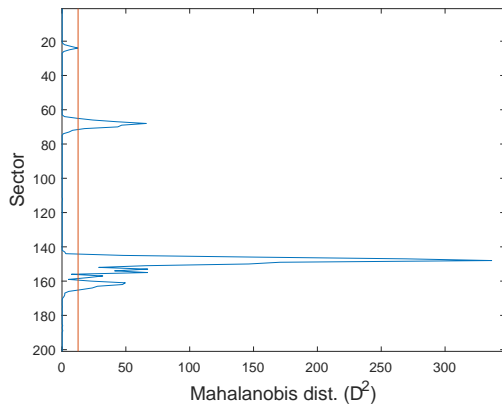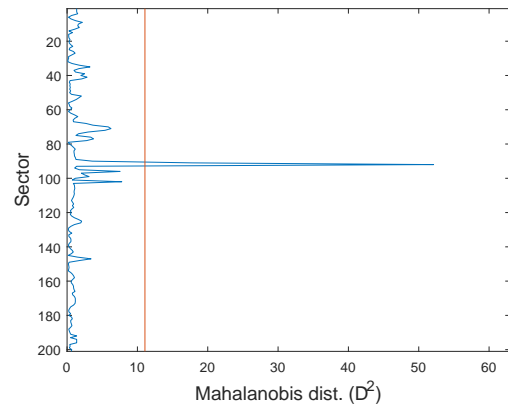
e) Full sector projections (5 PCs)   f) Full sector projections (15 PCs)
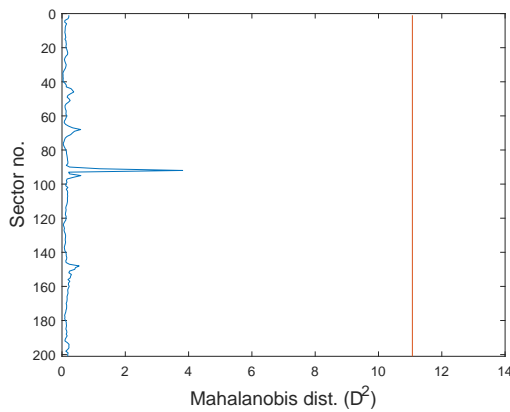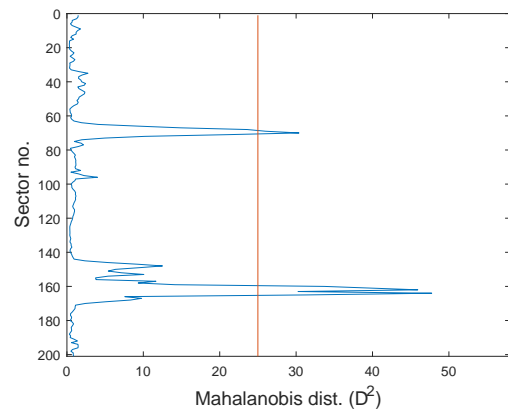
Figure 6.8: Mahalanobis distances for TB1PA1

111

a) Full sector projections

b) Constant A-scan projection (A-scan 66)
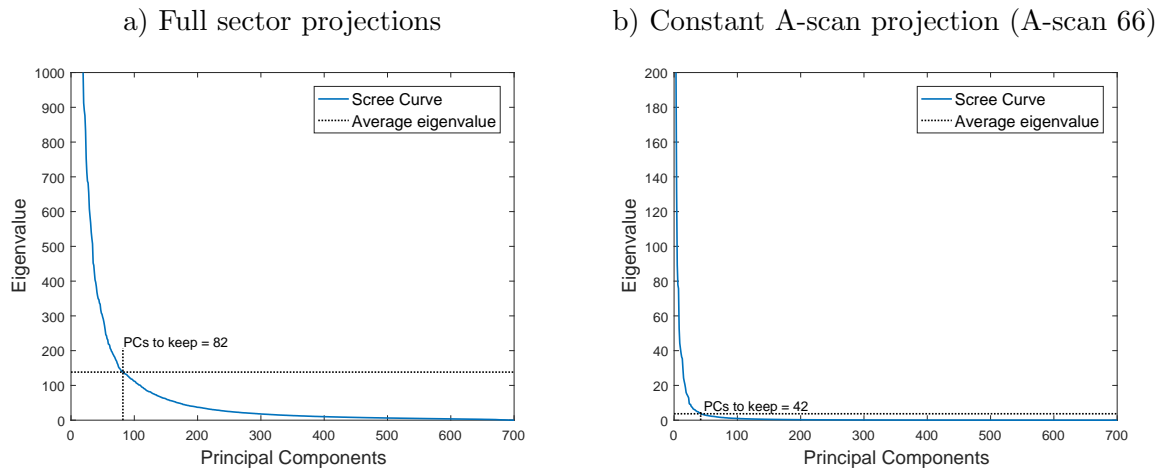
Figure 6.9: Scree plots 700 samples

a) Full sector projections (PCs = 82)

b) A-scan 66 (PCs = 42)
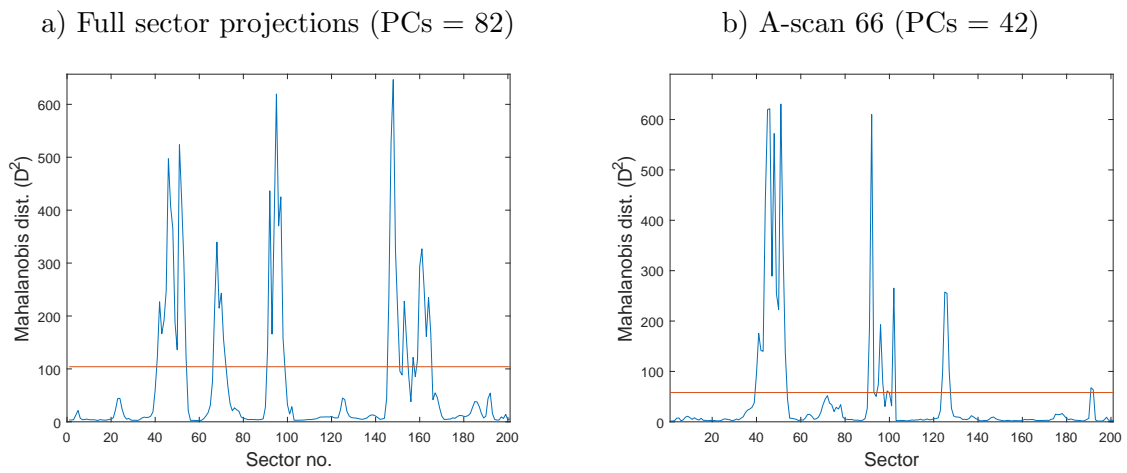
Figure 6.10: Mahalanobis distances (TB1PA1) with full training set

112

6.21 The Kaiser stopping rule

Although obtained from the scree plot, and using a common practice (section 6.11), establishing how many PCs to retain (the so called 'stopping rule') is a matter of some discussion [92]. It is claimed that the 'elbow' or 'hockey stick' method has a tendency to underestimate the dimension of the data [93]. Other authors suggest alternatives. In particular Ledesma and Valero-Mora [94] suggest a method due to Kaiser [95]. This simply takes the average value of all eigenvalues. The number of eigenvalues greater than the average is now the number of PCs to use.

Examples of this are given in figure 6.9. The 2 scree charts use all 700 observations of the available training set data. The first figure 6.9.a is for full sector projections. Application of the Kaiser stopping rule suggests using the first 82 PCs. Figure 6.9.b is an example for constant A-scans. It suggests using the first 42 PCs for A-scan 66; the chart for A-scan 35 (not illustrated) indicates 41 PCs.

Using the full training set and number of PCs from the Kaiser rule, figure 6.10.a illustrates the resulting Mahalanobis distances for full sector projections of TB1PA1. Similarly figure 6.10.b provides the result for a constant A-scan projection. In comparison with the earlier results, for the smaller training set (figure 6.8), these represent a considerable improvement. In particular, the constant A-scan orientation gives assertive indication of all anomalies intercepted by A-scan 66. In the case of the full sector projections there are no false indications but 2 faults are undetected. A reduction in the threshold level could improve performance, albeit at the potential expense of recording a false positive. At present selection of the threshold value is based on a representative $\chi^2_{(0.95,pc)}$ confidence level. Reducing the threshold requires a reduction in the confidence level. For reasons of comparison this value is maintained throughout this work.

These examples demonstrate that the training observations and number of PCs have some significant influence on the ability of the technique to discriminate between anomalous and anomaly free sectors. An increase in the number of training observations improves the estimate of the mean value of each dimension and the covariance matrix. However this is itself limited by the quality of the training observations. For example figure 6.7 indicates that some training examples are probably a closer estimate to the population than others. Removal of the more 'outlying' observations will likely lead to an improved estimate of the fault free condition. Before addressing this issue this chapter continues by further investigating the method using the present training set.

6.22 Evaluation using the manual training set

To provide a quantitative evaluation of this approach it is necessary to make use of the ground truth data, section 4.4. This is available in two forms. First there is a numeric list of faulty sectors and A-scans and secondly a binary mask. Full sector projections are only capable of highlighting sectors that potentially contain one, or more, anomalies; this is referred to as sector only look-up. Constant A-scan projections, when using a full set of A-scans, provide sector only look-up or sector and A-scan look-up of anomalies.

The following makes use of the ground truth test vectors and a test bench to automate an evaluation of each approach.

6.23 Full sector examples

Using the manually selected training data, the Kaiser stopping rule produces a value of 82 principal components, figure 6.11.a. Any full sector projection with a Mahalanobis distance greater then the $\chi^2_{(0.95,82)}$ confidence limit is considered to contain a fault. If this sector is listed in the test vectors then a True Positive (TP) is counted. If it is not listed a false positive (FP) is counted. Similarly a true negative (TN) or false negative (FN) is counted for a Mahalanobis distance below or equal to the confidence limit. Following a sequential scan of all sectors the accuracy of the technique is determined in the usual way (equation: 5.5) as are other quantities such as specificity (equation: 5.7) and sensitivity (equation: 5.6).

Table 6.1 lists the percentage accuracy results after comparing each test block with its respective set of test vectors. These are generally poor with the sensitivity figure indicating a low hit rate. Using a sweep of thresholds, starting from zero, the ROC results suggest that with a lower threshold greater accuracy with much improved sensitivity is possible, the corresponding reduction in specificity being slight. In addition figure 6.11.b indicates clear evidence of the 4 larger anomalies. There is some evidence of the two smaller pores but these are not distinct and are below the confidence level.

One method of improving the performance of full sector projections might be to increase the number of PCs. This is likely to have the effect of increasing the prominence of the two smaller anomalies. Experiments might lead to an alternative method of obtaining a better number of PCs to use. For these trials, which largely intend to demonstrate the method and obtain indicative methods of performance for future comparisons, this is not

a) Scree chart b) TB1PA1 using 82 Pcs
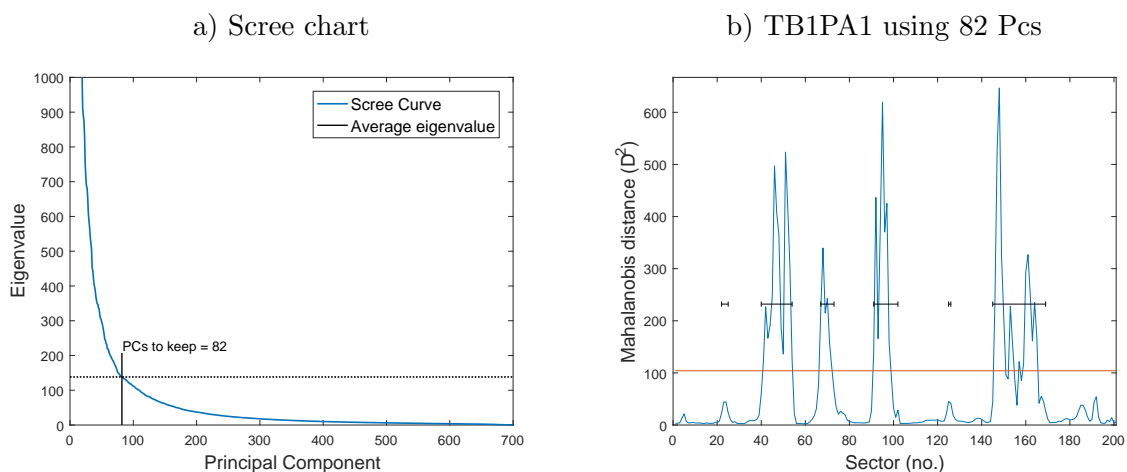
Figure 6.11: Examples of full sector projections

attempted. A present consideration is that the Kaiser rule gives a reasonable number of PCs and, more importantly, provides an objective method for their calculation.

Test Piece	χ^2				ROC				
	Acc (%)	Sens.	Spec.	F_{score}	AUC	Acc (%)	Sens.	Spec.	F_{score}
TB1PA1	89.1	0.66	0.99	0.80	0.99	93.5	0.98	0.91	0.86
TB1PA2	82.7	0.61	0.98	0.74	0.95	90.9	0.87	0.93	0.82
TB1PA3	86.6	0.59	0.99	0.75	0.99	95.2	0.99	0.94	0.85
TB1PA4	89.3	0.66	0.98	0.79	0.99	94.9	0.93	0.96	0.86
TB2PA1	88.2	0.62	0.99	0.76	0.98	92.7	0.97	0.91	0.83
TB2PA2	88.2	0.68	0.99	0.78	0.97	94.1	0.93	0.95	0.87
TB2PA3	82.4	0.47	0.99	0.64	0.95	92.2	0.93	0.92	0.78
TB2PA4	93.2	0.76	0.99	0.86	0.99	97.6	0.98	0.97	0.91
Averages.	87.5	0.63	0.99	0.76	0.98	94.0	0.95	0.94	0.85

Table 6.1: Accuracy (%) for full sector projections (p=82)

Test Piece	Thresholds %DR	
	$\chi^2_{(0.95,82)}$	ROC
TB1PA1	16.1	3.56
TB1PA2	15.2	5.85
TB1PA3	11.8	2.83
TB1PA4	14.7	4.92
TB2PA1	16.0	2.62
TB2PA2	12.5	4.80
TB2PA3	18.9	2.91
TB2PA4	16.0	7.37

Table 6.2: χ^2 and ROC thresholds as a % of dynamic range

6.24 Constant A-Scan examples

The introduction to the constant A-scan technique used only two example angles. For a more complete evaluation it is necessary to use the procedure for all A-scans. With the current test these are A-scans 1 to 71 covering angles $40°$ to $75°$ respectively (figure 4.7). For a complete test the procedure has to be repeated for each angle. With the current test data this requires 71 training sets with 71 repetitions of the test. Finally some method of aggregating the results of each projection for each sector is required. Unlike the full sector approach, which only identifies sectors containing an anomaly, this method identifies both the anomalous sectors and A-scans.

This section now provides initial results for constant angle A-scan projections over all angles. Using the same 700 manually selected observations a training, or reference, set is created for each angle; that is 71 training sets each with 1544 dimensions and 700 observations. Each training set has a different population of eigenvalues so that a different number of PCs is used for each projection. This also means that for each A-scan there is a different $\chi^2_{(0.95,pc)}$ confidence level. These are plotted in figure 6.12.a.

Recognition now consists of sequencing through the full set of test block sectors and projecting each A-scan against its respective training set. After each sequence of full sectors it is possible to plot the Mahalanobis distance for the A-scan (e.g. as in figure 6.10). Figure 6.12.b summarises the result of subtracting each threshold value from its respective A-scan projection. All negative values are now below the threshold so that any remaining positive distance is indicative of an anomaly. There are now two methods of detecting anomalies. These were previously referred to as sector only look-up and full-image look-up. For the constant A-scan orientation each is defined as follows:-

1. Sector only look-up summarises the results by taking only the maximum A-scan distance for each sector, as in figure 6.12.c. This is in fact a summary of sectors containing faults as obtained from the previous full sector approach. In a practical situation this approach provides only sector information. In test applications such as this it provides a fast method of comparing the results against the ground truth.

2. Full image look-up converts the three dimensional data (figure 6.12.b) into a binary image, figure 6.12.d. In practice these indicate areas of potential anomaly in terms of sector and A-scan location. The binary image provides input to the test bench for comparison with the corresponding ground truth map.

a) Kaiser PCs for each Ascan

b) D^2 Projections above confidence level

c) Max D^2 Projections across each A-scan

d) Map of potential anomalies
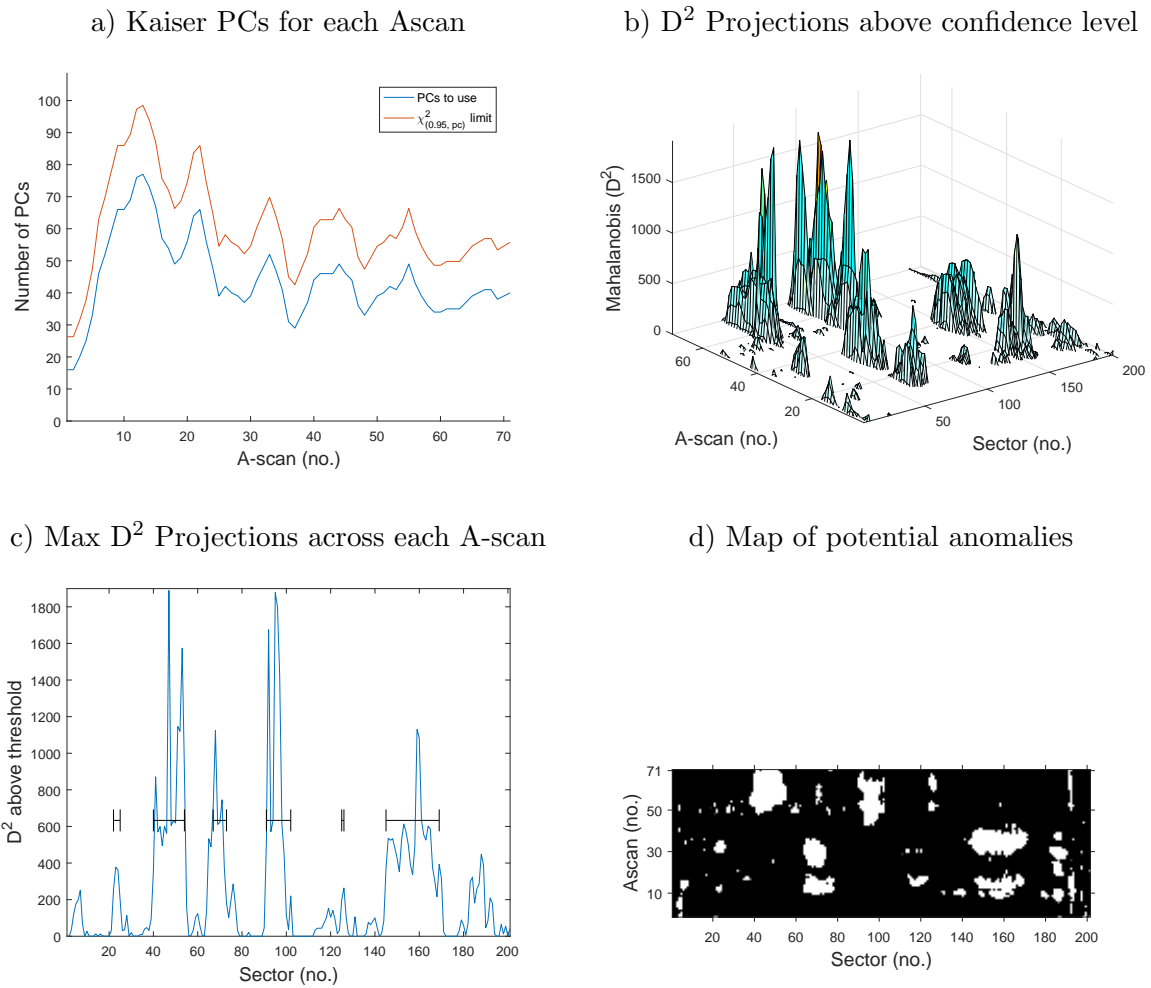
Figure 6.12: Results summary, A-scan training sets

6.24.1 Sector only look-up

Figure 6.12.c provides an illustration of the maximum Mahalanobis distance for each sector. Although it visually provides an indication of many of the anomalies it also contains many false positives leading to poor accuracy, table 6.3. Consequently the approach is inferior to that for full sector projections (table 6.1). Although the ROC analysis suggests a significant improvement to specificity is possible, it is difficult to automatically establish an overall threshold level to achieve this. The technique is likely to remain inferior to that for full image look-up.

6.24.2 Sector and A-scan look-up

This look-up approach contains most information. In essence a 2D binary version of the D^2 distances greater than the χ^2 confidence limit is compared with the binary ground truth mask. The results, table 6.4, are in strong contrast to those for sector only look-up and are comparable to those for full sector projections.

The ROC analysis demonstrates that, with a higher threshold, better classification is possible. It does not, however, suggest how this is achieved. One approach is to subject the resulting image data (from figure 6.12.b) to additional thresholding using the techniques introduced in chapter 5. An example for TB1PA1 is illustrated in figure 6.13. Table 6.5 lists the results after each additional threshold whilst table 6.6 lists the threshold values, as a percentage of the respective dynamic range.

The Otsu and ME methods produce thresholds that are consistently higher than those produced by the ROC analysis (table 6.6) whilst those given by the KI technique are much closer. It is also noticeable that, after subtracting the KI value from the original projection distances the accuracy, sensitivity and specificity values are in keeping with the 'optimum' values suggested by the ROC analysis.

Although additional Otsu and ME thresholding indicate the higher accuracies (table 6.5) it is also noticeable that the sensitivities are poor. This provides an example of accuracy alone when used for such comparisons. For this type of application, where a high probability of detection is essential, additional Otsu and ME thresholding is counterproductive. In the previous chapter, which dealt with the thresholding of sectorial images based on the peak A-scan values, the KI method also produced superior results for sensitivity (table 5.4).

Test Piece	χ^2				ROC				
	Acc (%)	Sens.	Spec.	F1	AUC	Acc (%)	Sens.	Spec.	F1
TB1PA1	55.2	1.0	0.34	0.61	0.94	87.4	0.88	0.86	0.85
TB1PA2	48.1	1.0	0.12	0.61	0.84	77.7	0.96	0.65	0.79
TB1PA3	62.68	1.0	0.44	0.65	0.92	85.7	0.93	0.82	0.85
TB1PA4	65.31	1.0	0.49	0.64	0.94	86.8	0.94	0.84	0.87
TB2PA1	52.45	1.0	0.32	0.64	0.91	81.8	0.95	0.76	0.81
TB2PA2	44.83	0.99	0.17	0.56	0.84	78.3	0.85	0.75	0.75
TB2PA3	66.34	1.0	0.49	0.65	0.92	87.1	0.96	0.82	0.83
TB2PA4	59.22	1.0	0.44	0.56	0.94	88.7	0.92	0.83	0.83

Table 6.3: Accuracy (%) for constant A-scan, sector only look-up

Test Piece	χ^2				ROC				
	Acc (%)	Sens.	Spec.	F1	AUC	Acc (%)	Sens.	Spec.	F1
TB1PA1	92.7	1.0	0.92	0.52	0.99	96.0	0.99	0.95	0.63
TB1PA2	85.8	1.0	0.85	0.37	0.98	93.2	0.96	0.93	0.56
TB1PA3	91.0	1.0	0.90	0.49	0.98	93.8	0.99	0.93	0.57
TB1PA4	91.7	0.98	0.91	0.47	0.98	93.2	0.98	0.93	0.52
TB2PA1	89.6	1.0	0.89	0.43	0.98	93.5	0.97	0.93	0.54
TB2PA2	89.3	0.97	0.89	0.50	0.96	93.6	0.93	0.94	0.62
TB2PA3	93.3	0.99	0.93	0.48	0.98	94.3	0.96	0.94	0.52
TB2PA4	93.3	1.0	0.93	0.49	0.99	96.4	0.99	0.96	0.64

Table 6.4: Accuracy (%) for constant A-scan, sector and A-scan (full image) look-up

a) Original D^2 above χ^2 limit

b) Ground truth

c) 'Optimum' ROC threshold

d) Kittler Illinworth threshold

d) Otsu threshold

e) Maximum Entropy threshold
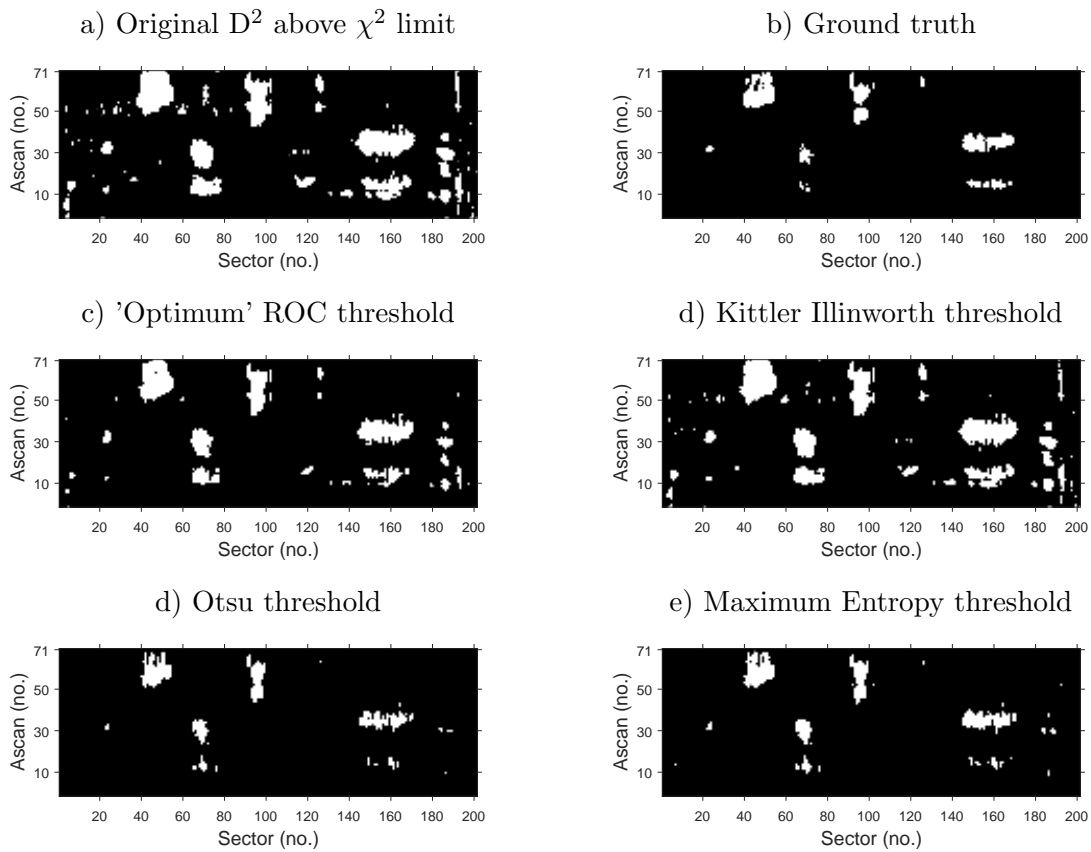
Figure 6.13: Results summary, A-scan training sets

For convenience table 6.7 provides a list of results for:-

- full image look-up using the χ^2 confidence level alone (from table 6.4)

- full image look-up using the χ^2 confidence level plus additional KI threshold (from table 6.5)

- KI thresholding of peak A-scan images (from table 5.4)

Test Piece	Kittler Illingworth			Otsu			Maximum Entropy		
	Acc (%)	Sens.	Spec.	Acc (%)	Sens.	Spec.	Acc (%)	Sens.	Spec.
TB1PA1	93.6	0.97	0.97	98.2	0.73	0.99	98.2	0.8	0.99
TB1PA2	88.9	0.99	0.88	96.8	0.64	0.98	96.1	0.14	0.99
TB1PA3	92.2	0.99	0.92	97.5	0.66	0.99	96.2	0.11	0.99
TB1PA4	92.5	0.99	0.92	97.5	0.72	0.98	97.5	0.16	0.99
TB2PA1	92.2	0.98	0.92	97.4	0.75	0.98	97.2	0.18	0.97
TB2PA2	91.0	0.95	0.91	96.3	0.68	0.98	95.0	0.13	0.99
TB2PA3	95.0	0.95	0.95	98.0	0.72	0.98	97.1	0.11	1.0
TB2PA4	96.5	0.99	0.96	97.5	0.21	1.0	97.7	0.27	0.99

Table 6.5: Constant A-scan look up with additional thresholding

Test Piece	ROC	KI	Otsu	ME
TB1PA1	0.5	0.6	10.6	8.23
TB1PA2	2.2	0.6	7.4	17.2
TB1PA3	1.5	0.6	8.6	23.9
TB1PA4	1.3	0.06	13.3	18.0
TB2PA1	2.3	0.06	11.4	9.4
TB2PA2	2.25	0.08	9.4	23.53
TB2PA3	0.3	0.3	18.63	10.6
TB2PA4	0.57	0.6	16.7	7.1

Table 6.6: Additional thresholds to χ^2 value (% of dynamic range)

Additional KI thresholding produces, in this case, the highest sensitivity. One cause of the overall low average sensitivity for KI thresholding of the peak A-scan only image is two particularly low values of 0.32 and 0.79. Otherwise thresholding of the A-scan peak image, which is computationally by far the most efficient, would be the preferred approach. On the basis of this evidence the χ^2 confidence level with, or without additional KI thresholding provides a more reliable classification. For this work the χ^2 confidence level is fixed at 95 %. If additional KI thresholding is to be used the probability of detection might improve by a slight reduction to the confidence level.

With this small number of test pieces it is not possibly to conclude, with certainty, that PCA has produced this slight improvement in performance. A further assessment is through a comparison of the SNRs.

6.25   SNR's from Mahalanobis distances

In the previous chapter anomaly detection was based on thresholding images created using the peak value of each A-scan. From the image it was possible to determine the SNR of

| Test Piece | Full image look-up | | | | | | peak A-scan image | | |
| | χ^2 | | | χ^2 + KI | | | KI only | | |
	Acc (%)	sens.	spec.	Acc (%)	sens.	spec.	Acc (%)	sens.	spec.
TB1PA1	92.7	1.0	0.92	93.6	0.97	0.97	95.5	1.0	0.95
TB1PA2	85.8	1.0	0.85	88.9	0.99	0.88	94.5	0.32	0.99
TB1PA3	91.0	1.0	0.90	92.2	0.99	0.92	94.9	1.0	0.95
TB1PA4	91.7	0.98	0.91	92.5	0.99	0.92	95.9	0.93	0.96
TB2PA1	89.6	1.0	0.89	92.2	0.98	0.92	96.1	0.97	0.96
TB2PA2	89.3	0.97	0.89	91.0	0.95	0.91	95.6	0.91	0.96
TB2PA3	93.3	0.99	0.93	95.0	0.95	0.95	95.6	0.96	0.96
TB2PA4	93.3	1.0	0.93	96.5	0.99	0.96	95.7	0.79	0.96
Averages	90.8	0.992	0.903	92.73	0.976	0.928	0.955	0.86	0.961

Table 6.7: Results - full image look-up with added KI thresholding and KI only

each anomaly, table 5.8. It is now possible to repeat this exercise using an image of the Mahalanobis distance of each projection. These results are presented in table 6.8. For convenience the previous results from the A-scan peaks are duplicated. In all cases it is found that the respective SNR is increased. This will contribute to an improvement in sensitivity.

| Fault Instance | SNRs (dB) | |
	A-scan Peaks	D^2
Pore 1	23.67	24.9
Crack	31.25	38.1
Porosity I1	24.22	34.7
Inclusion 1	29.58	38.5
Pore 2	18.78	33.1
Inclusion 2 I1	35.82	54.4
Porosity I2	18.54	29.9
Inclusion2 I2	25.67	29.5

Table 6.8: SNR's comparison A-scan peaks

This section demonstrates that the PCA projections are capable of visually highlighting areas containing potential anomalies. At present, the reference or training set is produced manually. For automatic anomaly detection this is not satisfactory. Thus not only are the results un-repeatable but there is no indication of whether there are sufficient original examples or whether there is a better alternative. These issues are dealt with in the following chapter.

6.26 TFM images

Attention is now given to TFM images. As in the case of sector scans the TFM data is in the form of a set of images at different index positions. Unlike a sector image, which is composed of A-scan samples, a TFM image (figure 6.14) is a rectangle of pixels with each intensity resulting from the TFM algorithm.
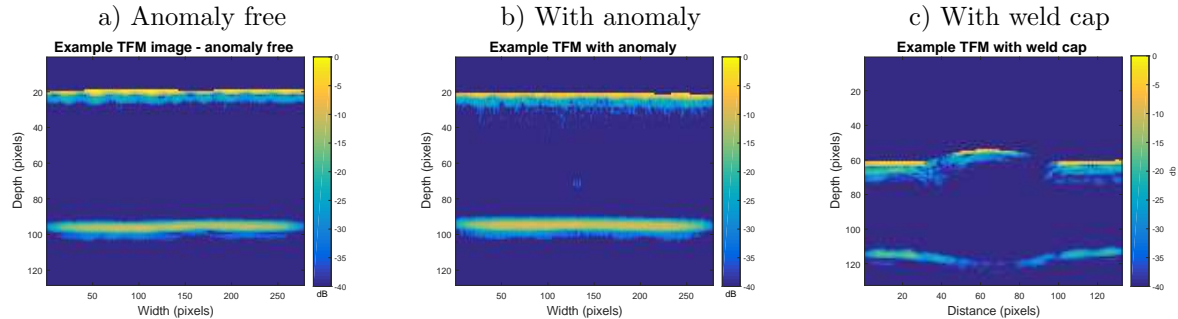
Figure 6.14: Example TFM images

For the images under consideration a further difference, with the sector scans, is that the data is from an immersion test at normal incidence rather than direct contact at oblique incidence. This, and the fact that each pixel is a single point of focus, means that the front wall interface becomes a dominant feature. In the case of oblique sector scans this interface is omitted due to gating. For exploratory purposes much of this initial work concerns the two, previously described, test pieces TFM1 and TFM2. These do not contain weld caps. The more irregular test piece, TFM3, which contains a weld cap and a relatively large side drilled hole, is considered after the exploratory studies. For convenience the co-ordinates on each TFM image scale are not converted to distance measurements, they are instead left as integer row and column numbers. In the test material each TFM pixel represents a 0.1 mm square. Overall dimensional details are found in the earlier work (figure 4.4).

The approach follows the approach for the sector images and the first part of the investigation is to consider how to arrange the data sets for PCA. The options are to consider each observation as a:-

- full image. This is equivalent to the previous full sector approach. Each observation is created by concatenating all of the columns of an image into a single vector.

- column of pixels at a constant offset from the centre of the weld, figure 6.15.a. Once again this is equivalent to the constant A-scan approach. Now a number of training sets are required at each offset location.

- row of pixels at a fixed depth, figure 6.15.b. The training set is now a number of such observations from fault free images. As in the case of the constant A-scan approach this requires a number of training sets corresponding to each depth.

a) Dimensions at constant offset   b) Dimensions at constant depth
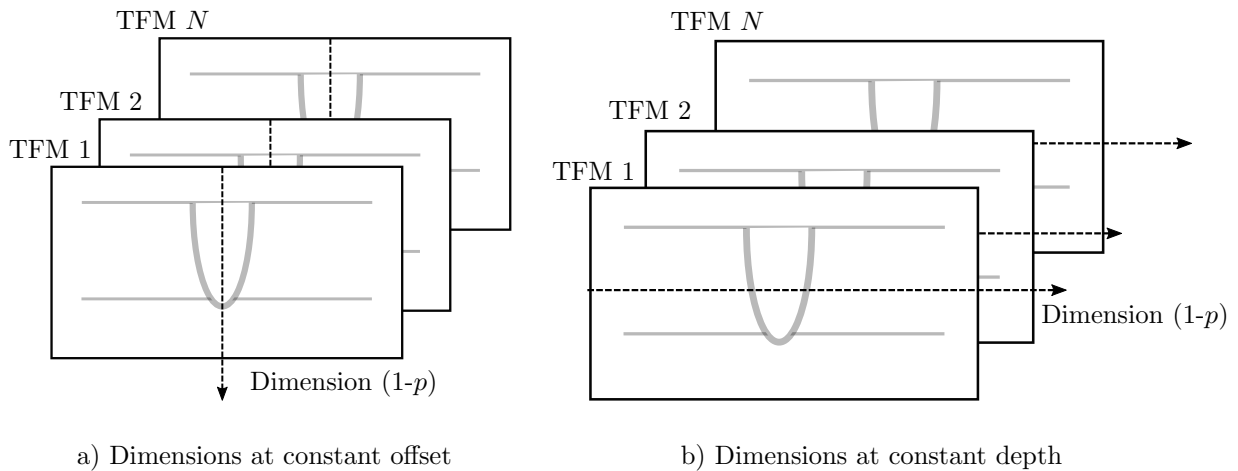
Figure 6.15: Orientations for TFM image analysis

Figure 6.16 presents a set of scree plots for TFM1. Using full TFM images, figures 6.16.a (anomaly free) and 6.16.b (with anomaly) are almost identical suggesting very little sensitivity to the presence or absence of an anomaly. In the next case, figures 6.16.c and d, the group of fixed columns (one anomaly free and one spanning an anomaly) are taken as the observations. In this case any difference between the two scree plots is slight and once again the approach is unlikely to be effective. In the last case (figures 6.16.e and f) the two groups are from a fixed row. Now there is a considerable difference between the two cases. In particular, for the anomaly, only 2 PCs are required to explain over 95% of the total variance; in the case of no anomaly less than 10% of the variance is explained. This is a much more encouraging result.

For the case of the TFM images the apparent lack of sensitivity of the first two approaches is likely to be due, in part, to the dominance of the front and back wall features (figure 6.14). When the PCA dimensions are vertical (through the front and back walls, figure 6.15.a) the front and back walls account for much of the variance regardless of the presence, or not, of an anomaly. For any row observation, taken approximately mid-way between the front and back wall, these dominant features are not present. Any anomaly that spans the selected row is now more distinct from the background.

a) Full TFM images anomaly free

TFM1 observatons - anomaly free (pxn) 35584x50, m = 49

b) Full TFM images with anomaly

TFM1 observations with anomaly (pxn) 35584x50, m = 49

c) Fixed column anomaly free

TFM1 (X index 139) anomaly free (pxn) 128x50, m = 49

d) Fixed column with anomaly

TFM1 (X index 139) with anomaly (pxn) 128x50, m = 49

(e) Fixed depth anomaly free

TFM1 (Z index 57) anomaly free (pxn) 278x50, m = 49

(f) Fixed depth with anomaly

TFM1 (Z Index 57) with anomaly (pxn) 278x50, m = 49

Figure 6.16: Example scree charts for TFM images

6.26.1 PCA projections of TFM images

Using each of the above orientations, figure 6.17 presents exploratory results of projections against respective training sets for each orientation. On all images the horizontal bars ('H' shaped) indicate the location of the artificial reflectors from the test vector file. In each case the stopping rule is that given by the Kaiser method. Similarly the red horizontal line, spanning the length of the x axis is the $\chi^2_{pc,0.95}$ confidence limit. If this line is not visible it is because its level is low compared with the peak distances. For the remaining examples in this chapter all TFM training sets use the same manually selected observations from the respective test piece. In the case of TFM1 and TFM2 each training set contains 200 TFM images. These are selected using the respective data sheets. In the case of TFM3, where images are taken using a finer indexing resolution (0.2 mm.) the training set contains 450 TFM images.

For full image projections anomalies are not identified. In fact, most projections lead to a distance above the χ^2 threshold and a consequent large number of false positives. For the constant offset examples the column of pixels is chosen to intercept all reflectors. Once again the results give little confidence in the ability of the approach to identify anomalies. These are in strong contrast to the last example, where the training set and observations are from constant depth projections. For TFM1 the reflectors are three flat bottomed holes drilled to the same depth with different diameters (figure 4.3). Each of these is clearly identified by the Mahalanobis distance peaks in figure 6.17.c. TFM2 contains 3 $\times\phi$1mm FBHs drilled to different depths (figure 4.3). The depth of the training set and observations correspond only to the deepest FBH which is clearly indicated (figure 6.17.f). The two leftmost FBHs are not at the detectable depth and as expected are missed.

6.26.2 Limitations of constant depth observations

In the examples of constant depth, figure 6.17.c, f and i, the training sets do not contain front and back walls. Consequently any anomaly at the observed depth represents an extreme deviation from the background and is readily detected by the major principal components. Where defects are in the body of the test piece and well separated from the front and back wall, this potentially provides a reliable method of detecting anomalies. However the approach is likely to fail for an anomaly located just under the surface of a weld cap. In such a situation the front wall will almost fill the observation (for example figure 6.14.c). In practice this limitation is further exacerbated by the fact that the front wall position is not flat across each indexed image. Differences are not limited to the offset
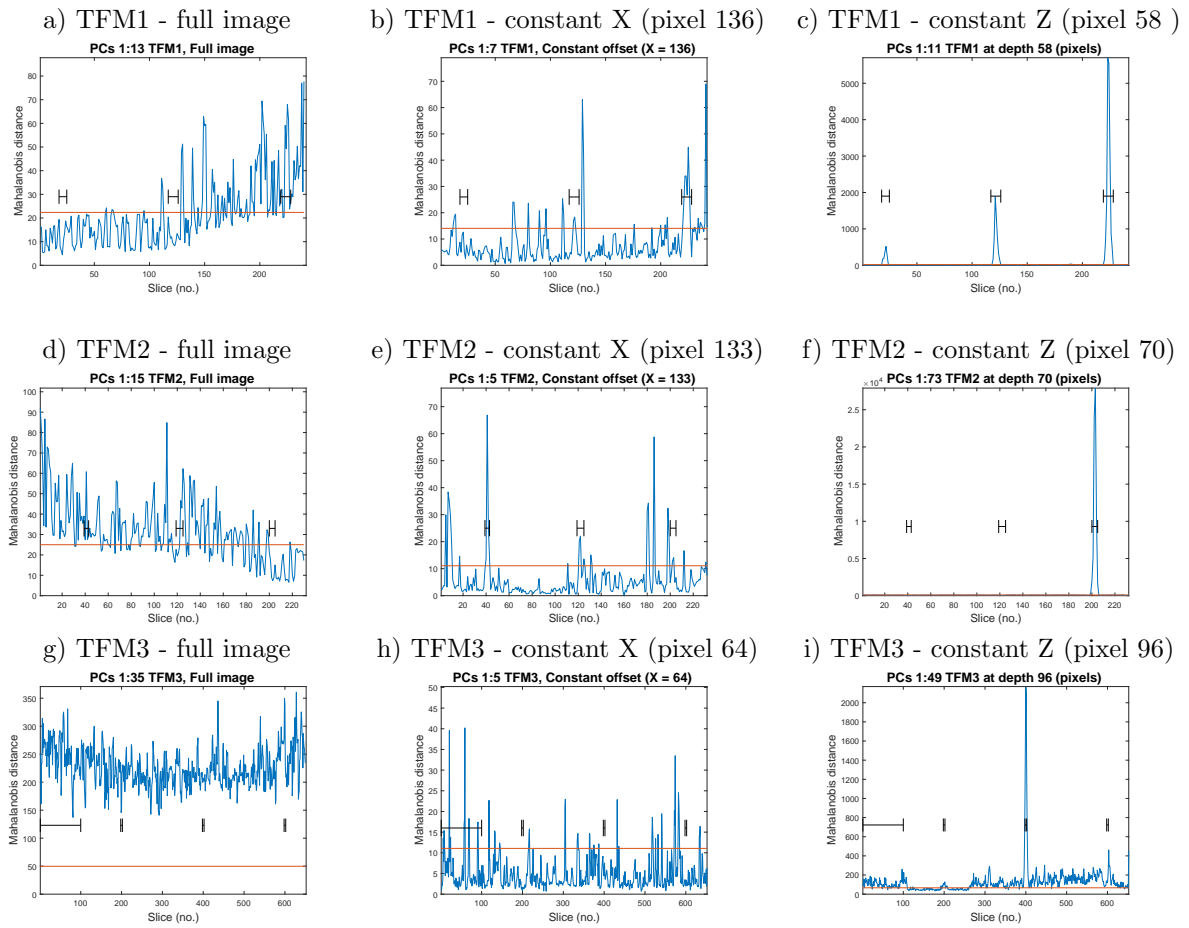
Figure 6.17: TFM - Anomaly detection at different orientations

126

position but are also evident as a skew where a slight rotation from image to image is evident, figure 6.18.

Even in a test tank it is not always physically possible to maintain a constant stand-off between probe and test piece over all index positions. Some further deviations may result from numerical accuracies within the TFM imaging algorithm itself. In a production environment pipe sections may not be perfectly aligned before welding and there is likely to be some eccentricity of the pipe relative to the probe. For both flat plate and curved surfaces, difficulties with alignment are further exacerbated by the process of welding itself which may cause some warping of the parent materials. In short, the likelihood that more than a few front wall surfaces will be alike is small. For PCA this issue is not likely to be so pronounced when the variables are taken as a column of vertical pixels as in the constant offset orientation. However the use of the constant depth orientation becomes questionable. There is now, in effect, no guarantee of a constant depth. Within the main body of the test material this is unlikely to be a major problem, adjacent depths being similar. However fluctuations in the presence and absence of front and back walls will cause difficulties for inspection at the inner surfaces. Consequently further evaluation will investigate only the full image and constant offset orientations.

a) TFM2 Slice 224 b) TFM2 Slice173
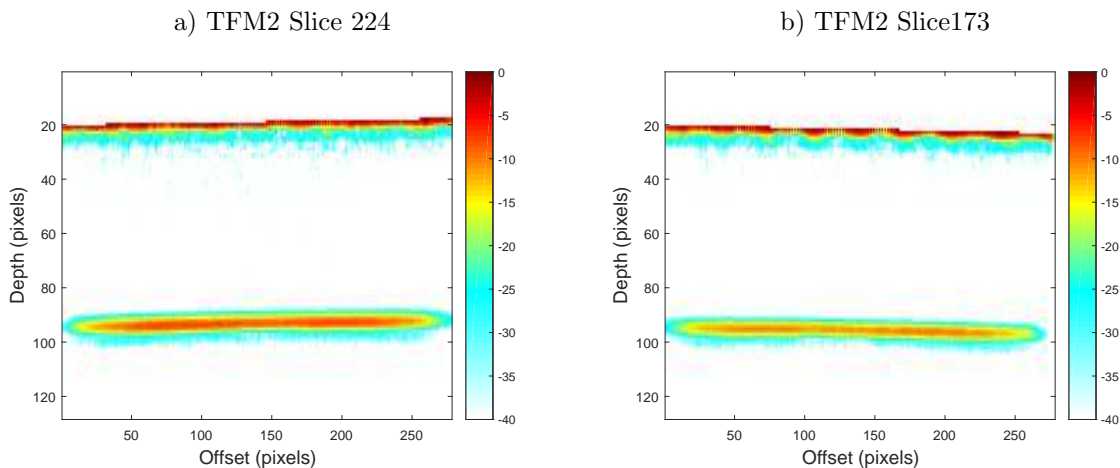
Figure 6.18: Example of image Skew

6.27 Anomaly enhancement

Previously PCA has been used in its conventional manner; that is, the major components are used to identify the direction of major variance in the data. A cursory look at the

original TFM images suggests that the strongest feature in each image is the front wall. Any variation in the front wall, such as its position due to skew or stand off, figure 6.18 is likely to account for much of the variance within the first few principal components. Variation due to anomalies are minor by comparison. This is confirmed by the previous projections. The following considers methods of addressing this balance with the intention of improving the probability of anomaly detection. To do this three methods are initially investigated, namely:-

- remove the front wall. Removing the feature responsible for the greatest variance will give greater prominence to the variance of smaller features. The front wall can contain evidence of critical surface breaking anomalies so its removal is not recommended. However at present this is the simplest way to deal with the potentially large source of variance.

- use a range of minor principal components. As demonstrated the major components are dominated by the main features. The less evident anomalies, by their nature, are small and do not contribute as greatly to the components of major variance. Anomalous conditions may be more evident within specific ranges of minor principal components.

- data scaling. Data scaling refers to the process of adjusting the standard deviation of each variable to a value of 1. With the variance of each variable, equally scaled anomalies will become, proportionally, more evident.

6.27.1   Front wall removal

As previously suggested (section 6.26.2) the location of the front wall from image to image is not always consistent and the number of pixels corresponding to its thickness also varies. This complicates front wall removal and introduces some subjectivity to the process.

To overcome these issues each column of the image is treated individually. Rows corresponding to the stand off, where there is no reflection, contain values of zero making them easy to identify. For each column the first set of rows containing zero, plus a number of rows representative of the front wall thickness, are deleted. The original length of the column is then re-instated by appending an appropriate number of zeros. For the investigation the thickness of the front wall, in terms of pixels, is decided manually and applied to all columns. The overall effect is to remove skew from the top surface and to bring all images into line. However for an original image with excessive skew, it may also lead to some

discontinuities, in the form of a zig-zag, across columns of the new image.

Figure 6.19 overviews the results of PCA projections for each test piece with the front wall removed. With the exception of the full image projection for TFM2, which remains useless, there is some considerable improvement over the respective images with the front wall intact (figures 6.19.a, c and e). This supports the assertion that the front wall does have a masking effect on anomaly detection. However front wall removal is not a viable solution to identifying small anomalies. Firstly subjectivity, in terms of deciding a thickness for the front wall, is not conducive to automation. More importantly, the front wall may contain evidence of critical surface breaking anomalies. In view of this and its masking effect, methods of mitigating its influence are needed.

6.27.2 Minor principal components

The previous work considers only the major principal components; that is, the components up to and including the number given by the Kaiser stopping rule. The remaining principal components ($p_{kaiser+1}$ to p_{max}) continue to represent linear functions of the original variables but with lower variance. Consequently any variation in these minor principal component scores, with respect to the covariance structure, is a potential indication of some smaller anomaly [90]. Based on this assertion Shyu et al. [82] propose a scheme called Principal Component Classification (PCC) which uses both the major and minor components. With PCC the major PCs detect extreme observations that have large values in some original features. However, the less commonly used, and not well documented, minor PCs are used to detect observations that do not conform to the normal covariance structure. The criteria for a potentially anomalous observation is now that it is either extreme in some original feature (as before) or that it does not conform to the expected covariance structure.

In an attempt to gain some understanding of the role of the minor PCs, for this application, the following presents the results of an investigation into their effect on the reconstruction of individual TFM images. In each case three SNRs are compared; namely that of the anomaly, a section of the front wall and a section of the back wall (tables 6.9 and 6.10). The area used for each measurement is indicated in figures 6.20.a (for TFM1) and 6.20.c (for TFM2). Selection of areas for the front and back walls is not entirely satisfactory. However, some experiments using different selections reveal that, although there is some variation in the measurements, the general trends remain consistent. For this reason these results are for guidance only; they offer some insight into the role of different ranges of PCs
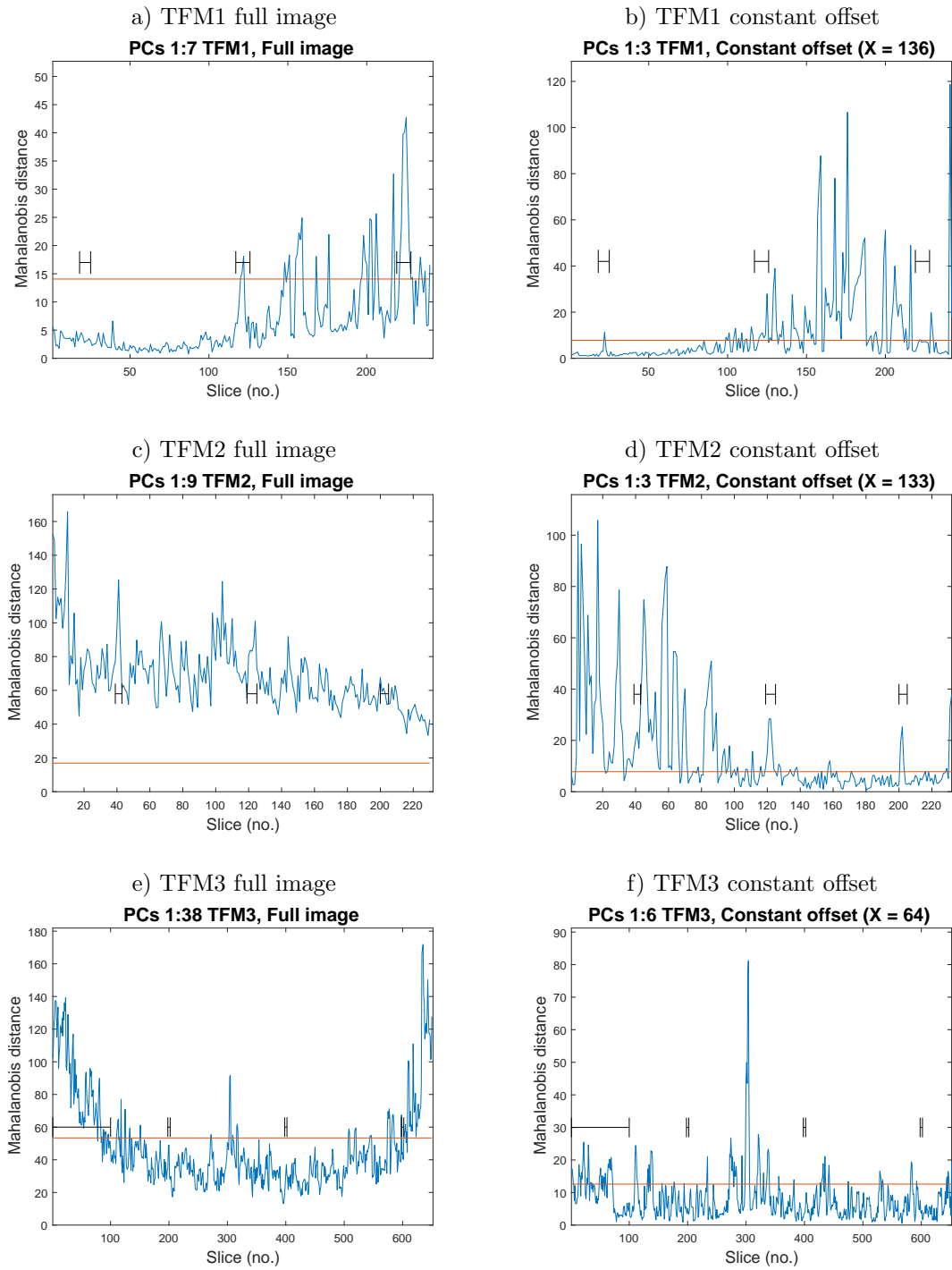
Figure 6.19: TFM major projections with front wall removed

rather than providing any conclusive recommendations.

Figure 6.20.b illustrates the TFM image for the ϕ3mm FBH. in TFM1 reconstructed from the first 5 PCs (Kaiser stopping rule). In this particular case the anomaly remains, in comparison to the front and back walls, quite prominent. Recovery from the major components rejects the components with lower variances, consequently improving the SNR of each feature, table 6.9. In this case a second application of the Kaiser stopping rule to the remaining PCs suggests a range of 6 to 17. As these contain smaller variances the SNRs reduce, almost back to the original values. Using all minor PCs (6 to 125) the SNR of the anomaly remains relatively unchanged whilst those for the front and back walls have reduced. For this large anomaly these results are not too significant other than to note, in comparison to the original image, recovery from all the minor PCs (6 to 125) has maintained the SNR of the large anomaly whilst reducing that for the two walls.

The second TFM image, figure 6.20.c, contains a smaller, ϕ1mm FBH, with a lower SNR. Once again a recovery using the major PCs results in some overall improvement, but in this case not so emphatic. As the anomaly is now significantly less prominent than the front and back walls there is some possibility that the second application of the Kaiser stopping rule will isolate its variance from that of the remaining background. Table 6.10 lists the SNRs for this second range (6 to 19) of PCs. It is significant that whilst the SNR for each wall reduces, that for the anomaly increases. Compared with the SNRs for the major PCs (1 to 5), those using all minor PCs (6 to 108) maintain the SNR for the FBH. whilst reducing those for the two walls.

A general conclusion from this investigation is that use of minor PCs has a tendency to improve the SNR (prominence) of anomalies whilst reducing that of the major features, in this case the front and back walls. Further investigations continue with a presentation of results using the minor PCs.

PC range	TFM1:224 SNRs (dB)		
	F wall	B. wall	Anomaly
original	51	47	26
1 to 5	59	51	39
6 to 17	53	46	25
1 to 17	58	51	25
6 to 125	47	40	24
17 to 125	47	39	25

Table 6.9: SNRs for TFM1, ϕ3mm FBH.

PC range	TFM2:206 SNRs (dB)		
	F wall	B. wall	Anomaly
original	42	39	16
1 to 5	45	36	21
6 to 19	34	29	26
1 to 19	41	34	25
6 to 108	33	25	22
19 to 108	35	25	18

Table 6.10: SNRs for TFM2 ϕ1mm FBH.
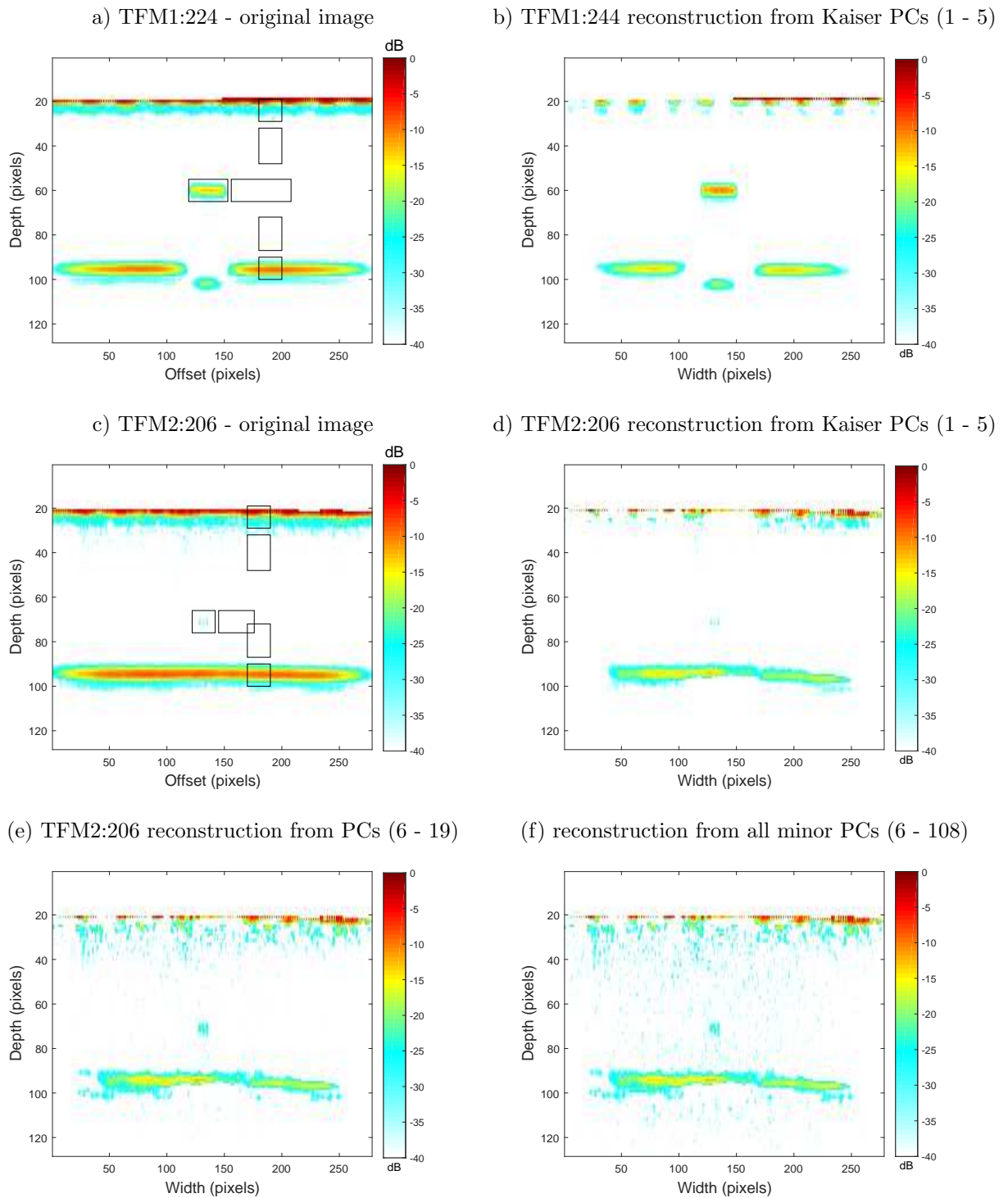
Figure 6.20: Comparison of PCA recovery from major and minor components

6.27.3 Examples using minor principal components

The tests of figures 6.17 are now repeated but this time using the minor principal component scores. In the case of constant depth observations there are no dominating front and back wall features and application of the minor principal components is unlikely to be of any benefit. This orientation is, therefore, omitted from further comparisons.

For the constant offset observations, projections using the full set of minor PCs is a considerable improvement on those using the major components, with both the front wall intact (figures 6.21.b, d and e) and removed (figures 6.19b, d and e). The explanation is that the major variances, due to the front and back wall features, are now omitted, the minor components reflecting variations due to smaller features such as anomalies.

For the full image projections Mahalanobis distances are, in all cases, well above the respective confidence threshold. In addition the characteristic pattern is for the plots to be relatively flat with some disturbance corresponding to each anomaly. For both TFM2 and TFM3 there are indications of each artificial anomaly. In the case of TFM1 the first anomaly (F1) is not distinct.

6.27.4 Examples using ranges of principal components

To complete this part of the investigation some consideration is given to a further decomposition of the minor principal components which are presently treated as a single block. Figure 6.20.e and f illustrate that in addition to giving more emphasis of any anomaly these also represent background noise. The rationale is that, whilst the major PCs separate variances due to the front and back walls and any anomaly, a second application of the Kaiser rule will separate variances due to anomaly and the background.

An overview of these results is given in figure 6.22. For full images all projections indicate a Mahalanobis distance greater than the threshold. For TFM1 the peak encompassing slice 224 corresponds to the test piece's strongest reflector (a ϕ3mm FBH). Similarly the distinct peak for TFM2 corresponds to its strongest reflector (the deepest ϕ1mm FBH around slice 40). In the case of TFM3 it is not possible to identify any reflector.

Results for constant offset projections indicate a degradation over the results using all minor components, although for TFM2 the anomalies remain distinct. Degradation is more severe for the full image projections where, with few exceptions, it is no longer possible to identify anomalies with any certainty.

a) TFM1 full image b) TFM1 constant X (pixel 136)

c) TFM2 full image d) TFM2 constant X (pixel 133)

d) TFM3 full image e) TFM3 constant X (pixel 64)
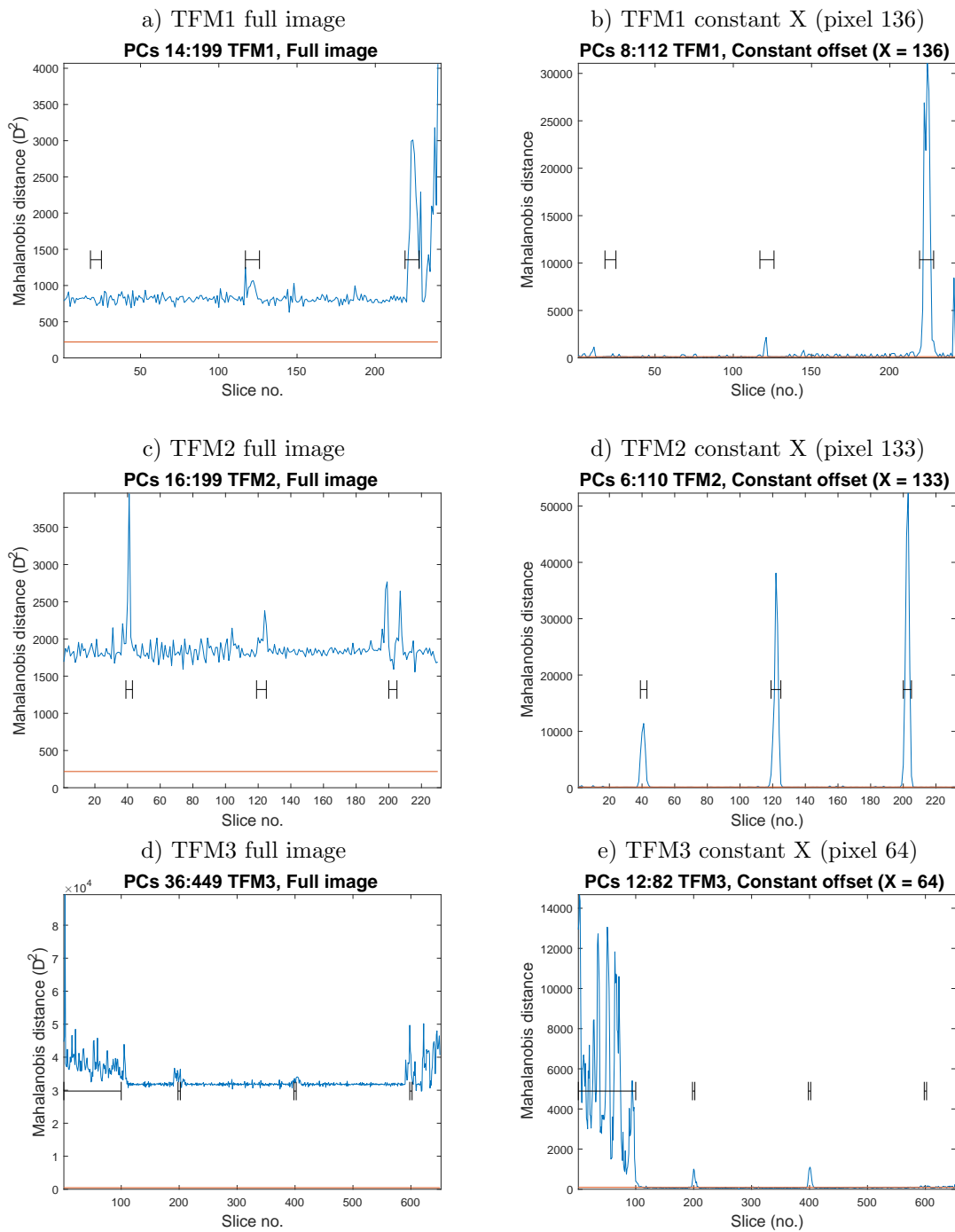
Figure 6.21: Anomaly detection using all minor components

a) TFM1 full image

b) TFM1 constant offset

c) TFM2 full image

d)TFM2 constant Offset

e) TFM3 full image

f) TFM3 constant offset
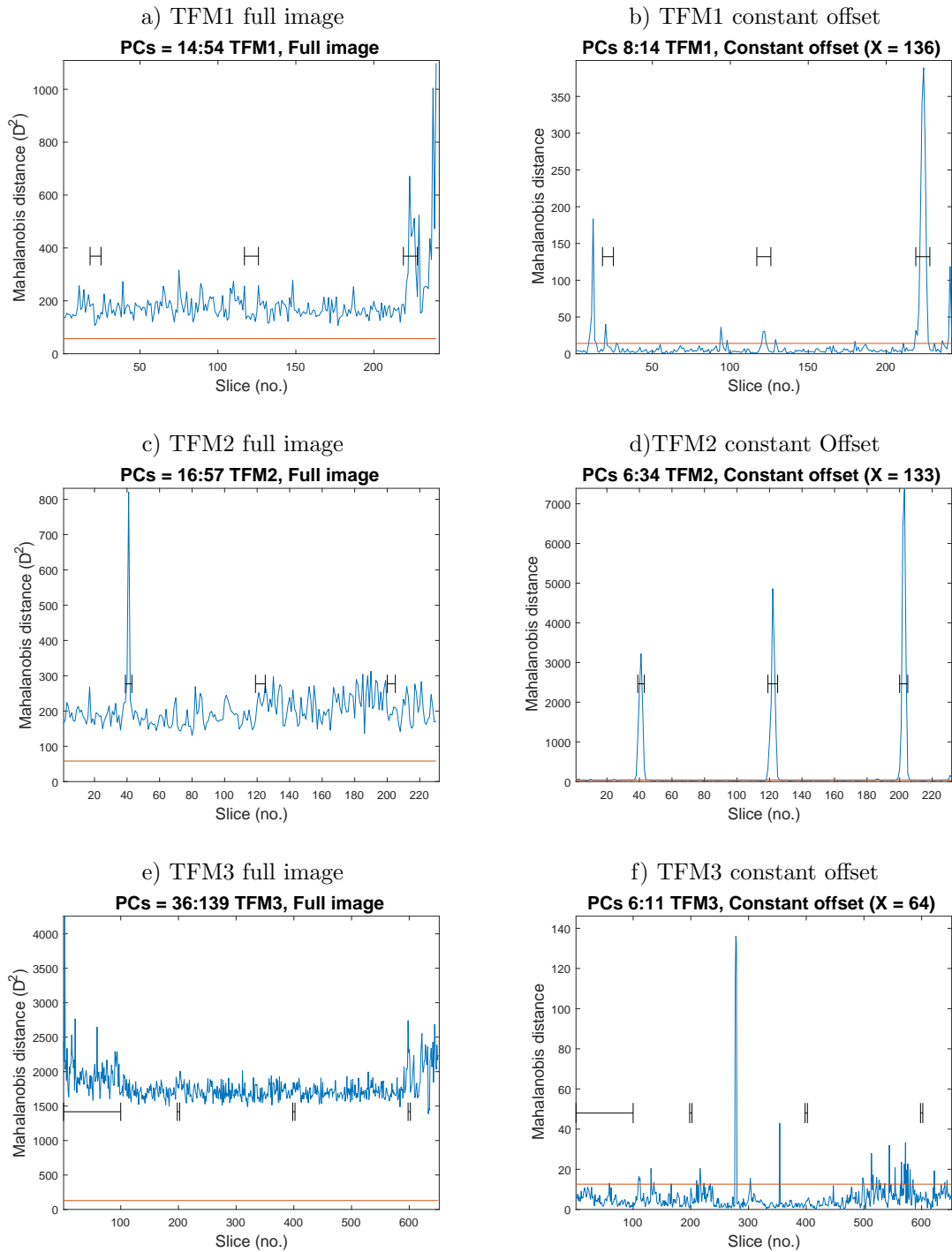
Figure 6.22: Anomaly detection using a range of minor PCs

6.27.5 Data standardisation

In all of the preceding examples data is from a single ultrasonic probe. Each element is considered identical and each channel is handled in exactly the same way. When each variable is measured on the same scale there is usually no need to account for the differences in the range of each. This is in contrast to applications outside ultrasonic testing where it is not uncommon for variables to be measured in different units and different scales (for example one variable may be in kilograms, another in centimetres etc.). In an attempt to assign greater effort to the measurements having smaller ranges than those of the front wall, this section investigates the application of data scaling. This is simple to apply. Apart from mean-centering no pre-processing has so far been done prior to constructing the covariance matrix. To standardise the data so that each variable has unit variance all that is required is for each mean-centred sample to be divided by the variable's standard deviation (σ):-

$$x' = \frac{x - \bar{x}}{\sigma} \tag{6.16}$$

After standardising, the PCA procedure continues as normal with the previous covariance matrix now referred to as the correlation matrix.

Figure 6.23 illustrates the result of data standardisation for each test piece using full sector projections and constant offset projections. Due to standardisation the projections use only the major principal components. Unit variance also means that, although the y-axis remains labelled as the Mahalanobis distance this is actually the Euclidean distance. Results for the full image projections of each TFM image (figure 6.23) provide no evidence of notable improvement. This is in strong contrast to those for the constant offset projections. Now all anomalies are clearly identified with very few false positives. In fact the result is comparable to that obtained through the projection of minor principal components (figure 6.21).

6.28 Review of orientations

With the available TFM images the front wall presents the greatest challenge for anomaly detection. This will apply to any data taken at normal incidence where it is not possible to gate out the front wall reflection without losing valuable data from the test piece surface. In the case of oblique incidence, such as the sector scan data, the front wall reflections are not recorded so that the data does not contain this dominant feature.

a) TFM1 full image

PCs 1:37 TFM1, Full image

b) TFM1 constant offset

PCs 1:33 TFM1, Constant offset (X = 136)

c) TFM2 full image

PCs 1:44 TFM2, Full image

d) TFM2 constant offset

PCs 1:33 TFM2, Constant offset (X = 133)

e) TFM3 full image

PCs 1:110 TFM3, Full image

f) TFM3 constant offset

PCs 1:26 TFM3, Constant offset (X = 61)
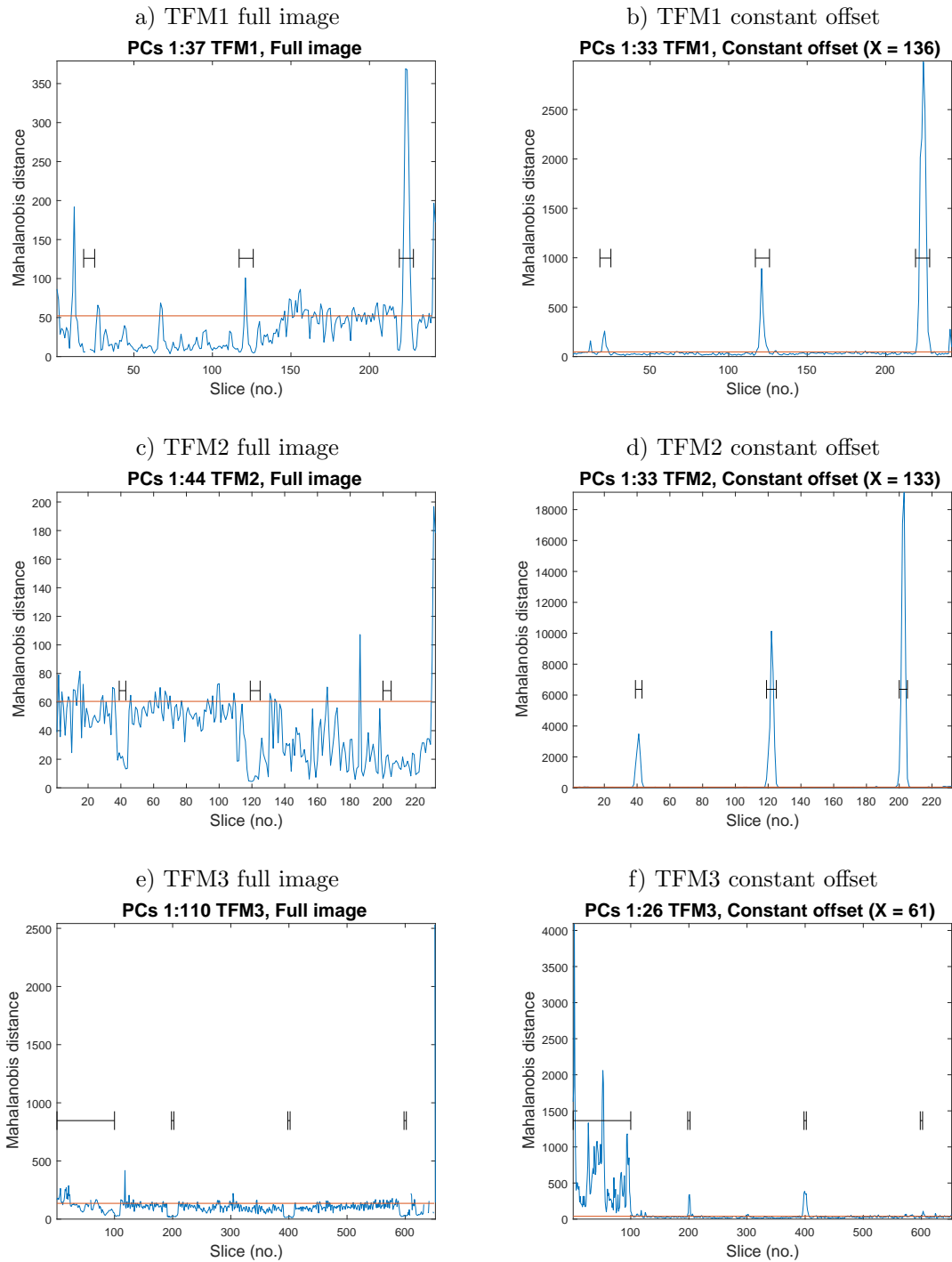
Figure 6.23: Major PC projections using scaled data

137

Attempts to remove the front wall were not satisfactory but did demonstrate that this feature masks anomalies. Of the three orientations considered, results from full image projections, where each image is converted to a single column vector, are not consistent. Further investigation might reveal a more reliable method of data analysis for this orientation. However results from constant offset projections appear, so far, to be far superior. In particular, with data standardisation, anomalies are pronounced using the major principal components. Although a number of training sets are required the dimensions are much reduced compared to full image projection. In practice it is likely that only the central offsets need to be analysed, those at the outer extremes, and far removed from critical regions, being omitted to improve processing efficiency. A further feature of the constant offset approach is the possibility of creating a 3-D visualisation of the test piece similar to that for the constant A-scan projections of the sector scans (figures 6.12).

Constant Depth observations offer a potentially reliable method of anomaly detection within the body of the specimen using the major principal components. Some uncertainty remains with the method as the depth approaches either that of the front or back wall. Once again methods of suppressing the front or back wall quickly become obfuscated. Where a weld cap is in place it is not possible to inspect regions higher than the parent material. Even in cases where the weld surface is dressed, the weld root is frequently out of the path.

For the case of an immersion probe mounted at normal incidence the above comments, substantiated by the results, provide a strong argument for directing the investigation towards the constant offset orientation. In addition the reference and test data is standardised so that the Mahalanobis distances are now equivalent to the Euclidean distances.

6.29 Constant offset projections with standardised data

Figures 6.23.b, d and e are created for a single offset. In each case the particular offset was chosen to intercept, as near as possible, each of the manufactured anomalies. Figure 6.24 illustrates the result of repeating this exercise for all offsets.

As in the case of the constant A-scan approach (figure 6.12.b) the $\chi^2_{pc}(0.95)$ threshold is subtracted from each offset before creating the image. This means that any value above zero represents a potential anomaly. Of particular interest is the fact that to create these images no explicit attempt has been made to remove the front wall. The plots are the result of data standardisation. Figure 6.26 represents the corresponding two dimensional images

a) TFM1   b) TFM2
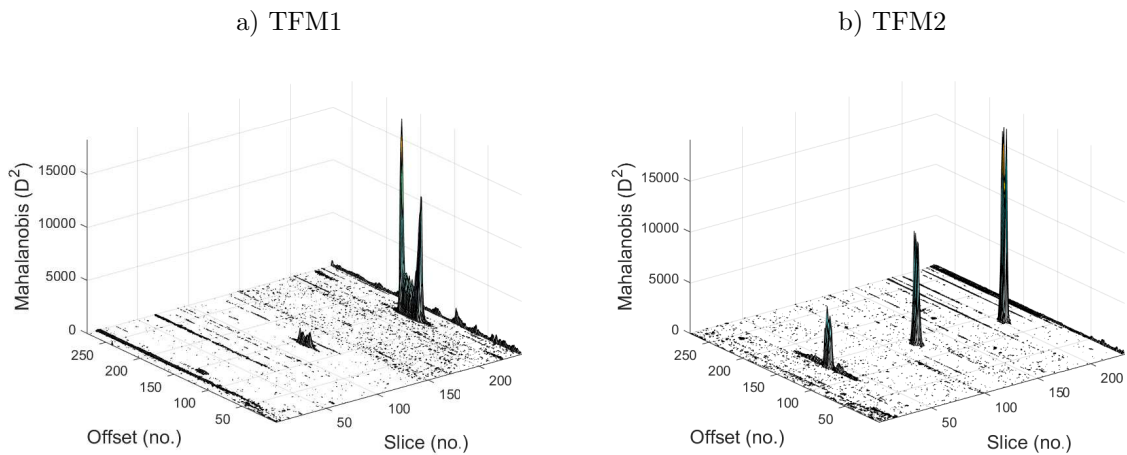
Figure 6.24: Full projections (above confidence limit) for constant offset orientation

a) TFM3   b) TFM3 Clipped
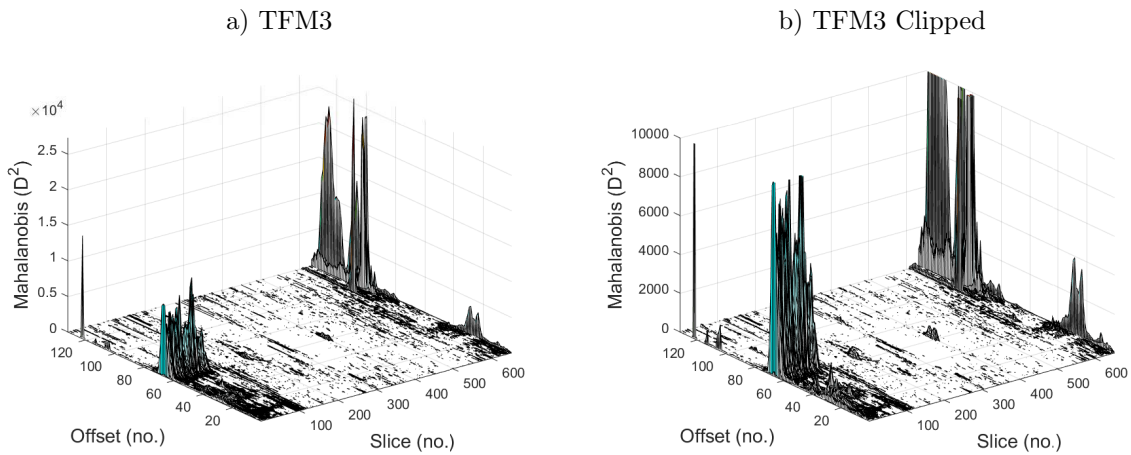
Figure 6.25: TFM3 full projections (above confidence limit) for constant offset orientation

for TFM1 and TFM2 using a db scale. Collectively both the three and two dimensional representations contain evidence of each anomaly; however some aspects of these indicators require explanation.

Variation in the relative magnitude of each reflector is the most notable feature of these images. This is explained by understanding that each projection distance relates to the total difference between observation and the reference set. Where the difference is large so too will be the projection. Figures 6.27 and 6.28, which illustrate TFM images of slices

139

corresponding to each FBH, indicate that differences are not due to the size or depth of the FBH alone. A significant contributory factor is variation in the back wall. Reasons for this are due to a combination of effects. The most significant fact is that under each FBH a back wall does not physically exist. However the location is also shadowed by the tip of the respective FBH. (as would be the case for a real internal anomaly). Where the back wall gap is less pronounced this is likely to be due to secondary reflections of the FBH, mode conversions and dispersion of the ultrasound.

The above gives some explanation for the characteristic peaks at either side of each FBH; most notably for the largest FBH on TFM1. In the case of TFM1's ϕ1mm FBH the back wall image is almost intact so that there is little variation and the projection is suppressed. This also applies to the deepest FBH for TFM2. In this case not only is the back wall image almost intact but it also contains higher than normal magnitudes either side of the FBH.

6.30   Weld caps

Unlike the two test pieces used in the previous test the third, TFM3, consists of two plates welded together. This is more representative of a real situation and is yet more complicated by the fact that neither the weld cap or root bead is dressed. The resulting irregularities of the front wall become a further difficulty in any attempt at its removal. Once again, however, the combination of all offsets over all sectors illustrated in figures 6.25.a and b suggests that with data standardisation the front and back walls become effectively transparent.

Figure 6.25.a is dominated by a number of projections in the end slices (around slice number 600). Although their corresponding offsets are away from the weld they remain dominant features. When the scale is clipped, as in figure 6.25.b, the side drilled hole and the three radial bottomed holes are more apparent. A 2D map of the peak D^2 distances, figure 6.29.a, illustrates a primary concern of these dominating features. For the anomalies to be visible the scale has been extended to -80 dB. With the scale down to -40 dB, as in figure 6.29.b, many of these features are not visible. This poor signal to noise ratio will hinder anomaly detection through comparison of a binary converted version of the image to the ground truth. In practice this could be mitigated by using only offsets that immediately surround the centre line of the weld (for example offsets corresponding to physical dimensions ± 5 mm. around the centre). However for the present, comparisons will be over the entire range of imaged offsets.

a) TFM1   b) TFM2
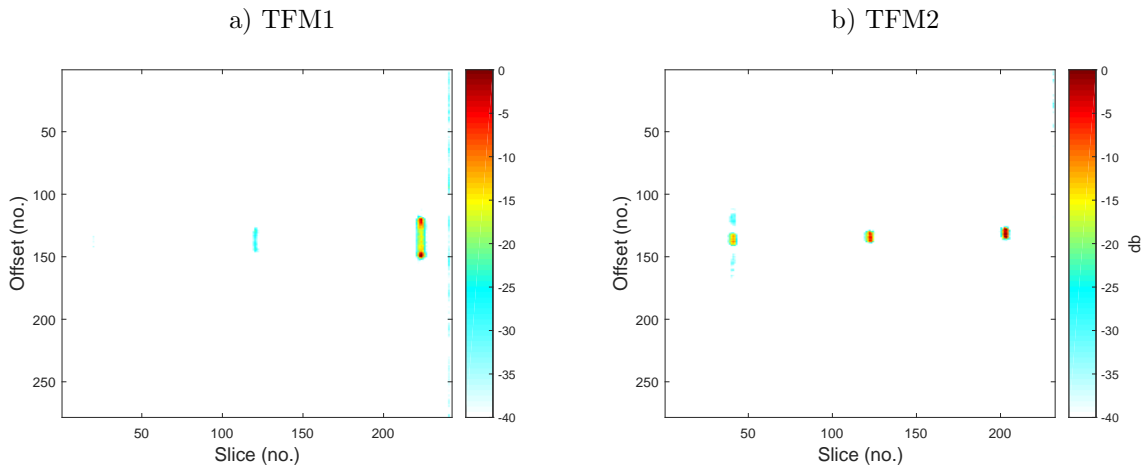
Figure 6.26: Maps of peak D^2 for TFM1 and TFM2

a) TFM1 - slice 21   b) TFM1 - slice 122   c) TFM1 - slice 224
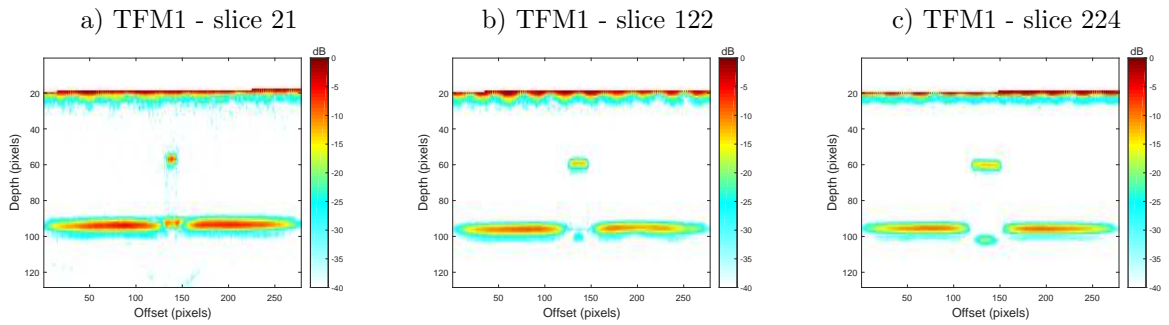
Figure 6.27: TFM1 - TFM images

a) TFM2 - slice 41   b) TFM2 - slice 122   c) TFM2 - slice 202
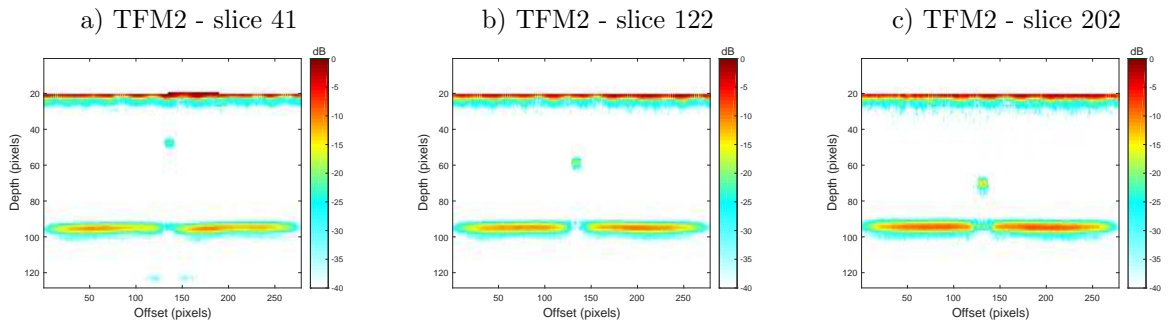
Figure 6.28: TFM2 - TFM images

At this point some features of figure 6.29.a are worth mention. Firstly there is evidence of the 3 SBHs and the SDH. Visual inspection of the test piece also reveals that the region marked as an unintended anomaly is actually present. After some estimations data relating to its size and location is included in the ground truth data. Finally two areas are

marked as 'persistent false positives'. These have not been included in the ground truth data. However to a varying extent they persist throughout much of the present and future analysis. Visual examination finds no evidence for their presence. They are thought to be caused by small inclusions as a result of the welding process. This cannot be confirmed without destruction of the test piece. They are not included in the ground truth and are therefore likely to contribute to the false positive count.

a) Down to -80 dB   b) Down to -40 dB
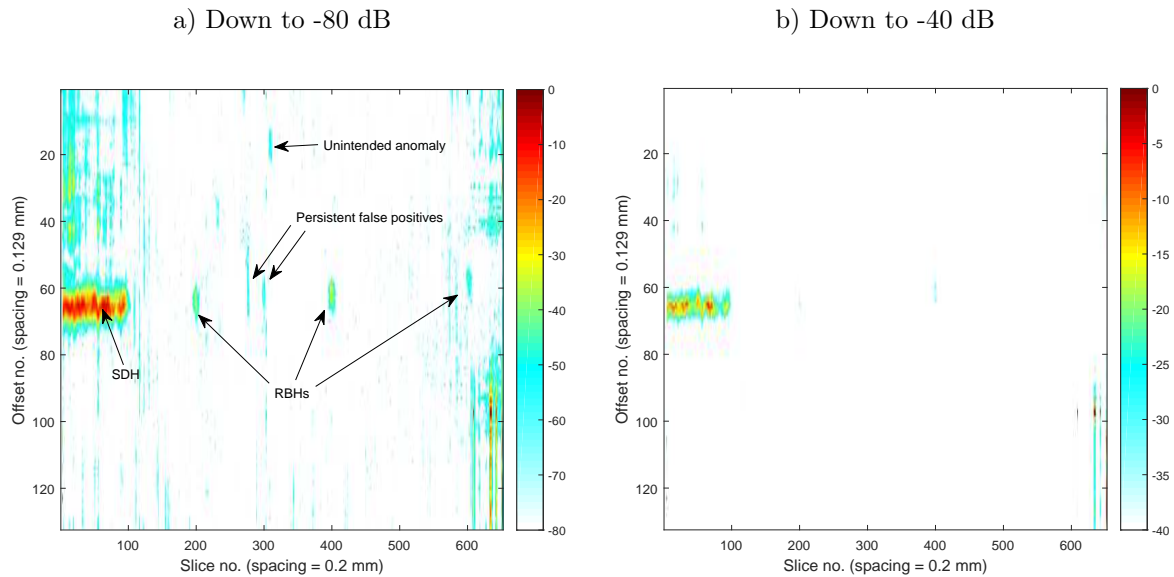
Figure 6.29: Map of peak D^2 for TFM3

6.31 Assessment of constant offset orientation

As in the case of the constant A-scan orientation, when considering sectorial data, the TFM constant offset orientation supports two methods of anomaly look-up. These are slice only look-up and slice with offset look-up. Slice only look-up takes the maximum distance (D^2) from all offsets belonging to the slice. If any of these are greater than the confidence limit then an anomaly is assumed. This provides an indication of slices containing potential anomalies. The alternative, slice and offset look-up, makes use of the entire set of projections. It locates potential anomalies by both slice and offsets. The following sections consider and evaluate each look-up method in turn.

6.31.1 Slice only look-up

Taking the maximum Mahalanobis distance across all offsets corresponding to each index position enables a plot of maximum D^2 across all slices to be made. Figure 6.30 provides a summary of these for each TFM image. Before producing the plots each offset has had its respective $\chi^2_{(pc,0.95)}$ confidence level subtracted. Now any value greater than zero represents a slice containing a potential anomaly. Visually these plots give confidence in the ability of slice only look-up to identify anomalies. Unfortunately, as a means of automatic anomaly detection, the confusion statistics listed in the χ^2 columns of table 6.11, do not confirm this. The poor results are due to many projections which, although small, nevertheless exceed the confidence level to give many false positives. This is confirmed by the ROC analysis which shows that a very small increase in threshold produces a considerable improvement in the both specificity and sensitivity. The associated F1 statistics give further confidence in the ROC analysis. However in the absence of a method of obtaining a reliable threshold these values remain potential only. For this reason slice only look-up is, at this stage, given no further consideration.
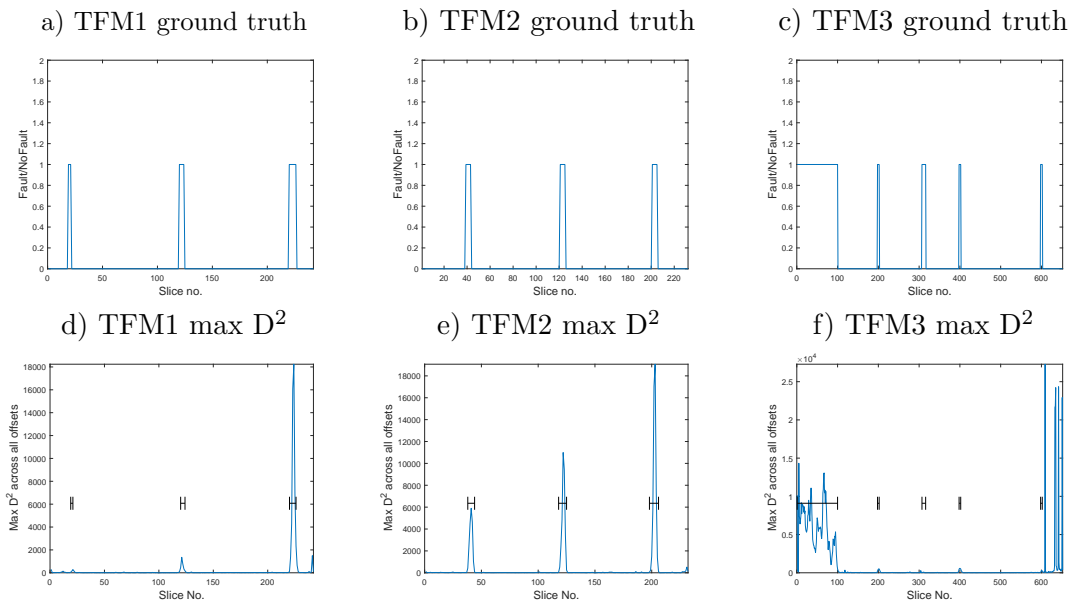
Figure 6.30: Slice only look-up

143

Test Piece	χ^2				ROC					
	Acc. (%)	Sens.	Spec.	F1	th (% dR.)	AUC	Acc. (%)	Sens.	Spec.	F1
TFM1	14.0	1.0	0.08	0.13	0.52	0.98	93.8	1.0	0.93	0.67
TFM2	9.1	1.0	0.03	0.12	1.4	0.96	97.8	1.0	0.98	0.86
TFM3	22.7	0.99	0.05	0.33	0.7	0.95	92.8	0.90	0.93	0.83

Table 6.11: Accuracy (%) for constant offset orientation - slice only look-up

6.31.2 Slice and offset look-up

Previous figures (6.24 and 6.25) illustrate the projections for each test using slice and offset look-up. As previously stated these projections already have the respective $\chi^2_{(0.95,pc)}$ confidence limit subtracted and they can readily be converted to a binary image for comparison with their respective ground truth image. Figure 6.31 illustrates a set of these binary images alongside their respective ground truths. A simple logical comparison between a binary image and its respective ground truth now provides a set of confusion statistics.

The χ^2 columns of table 6.12 lists the results of each comparison in terms of classification accuracy, sensitivity and specificity. Overall these results are a considerable improvement over those for slice only look-up. Of some significance is the fact that in each χ^2 image there is clear visual evidence of all anomalies.

The next stage is to establish if improvements are possible. Table 6.12 lists the results of the ROC analysis using the χ^2 projections. In all cases the AUC indicates the potential for classification. The fact that there are only slight differences between sensitivities and specificities (for χ^2 and ROC) is evidence that the χ^2 confidence level already provides a reliable classification. The poor F1 statistics are more representative of imbalances in the data set rather than the behaviour of the ROC analysis. Evidence of this is represented by the actual ROC curve (figure 6.32), the high specificity values, and the images created using the respective ROC thresholds (figures 6.34.c, 6.35.c and 6.36.c).

Test Piece	χ^2				ROC					
	Acc (%)	Sens.	Spec.	F1	$\Delta R_{th\%dR.}$	AUC	Acc (%)	Sens.	Spec.	F1
TFM1	92.0	1.0	0.92	0.07	0.005	0.99	98.5	1.00	0.98	0.28
TFM2	92.0	0.94	0.92	0.03	0.003	0.96	95.2	0.94	0.95	0.05
TFM3	85.5	0.94	0.85	0.12	0.002	0.96	94.5	0.92	0.95	0.26

Table 6.12: Comparison of χ^2 and ROC thresholds

Based on a tradeoff between sensitivity and specificity the ROC thresholds are only available with a ground truth. In practice the χ^2 confidence level may be used as a threshold to produce images such as those in figure 6.31. Alternatively the D^2 projections can be

144
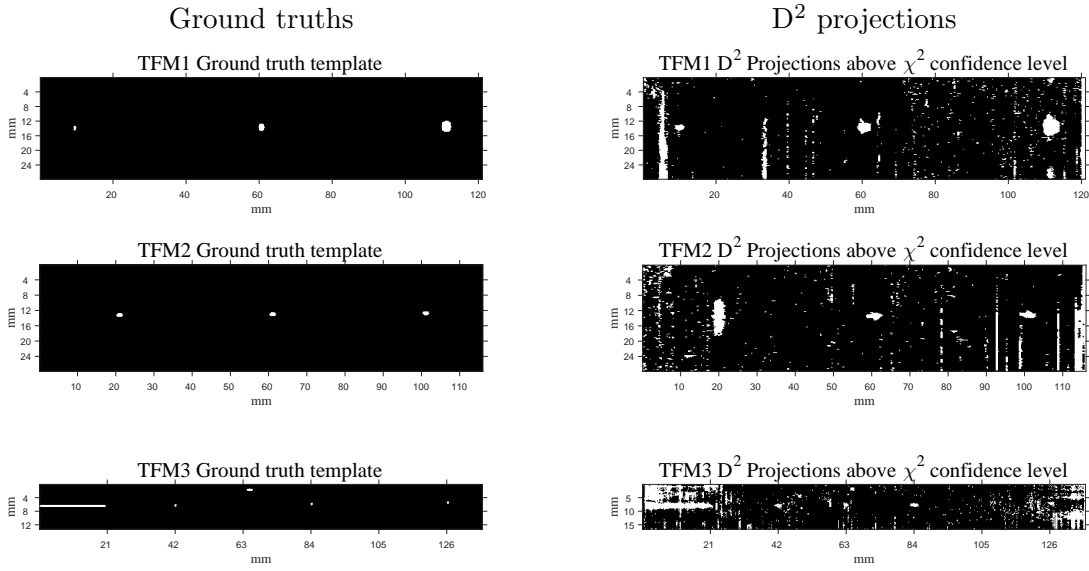
Figure 6.31: Ground truths and D^2 projections for TFM1, 2 and 3
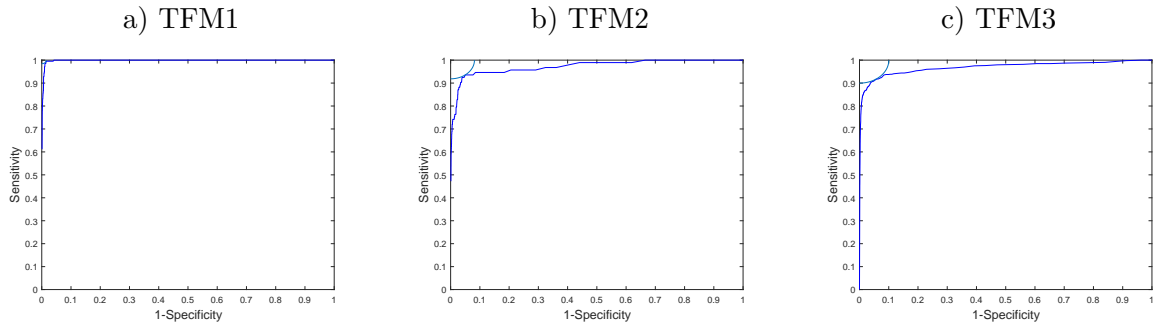
Figure 6.32: Example ROC curves

subject to one of the previously introduced image thresholding techniques. It became apparent that expressing such comparisons as a percentage of the total dynamic range of the projections was not satisfactory. This is due primarily to the fact that projections can contain spuriously high values. This is mostly evident in the case of TFM3 (figures 6.30.f and 6.25. In later work there are instances where this is exacerbated. It was therefore decided to use an alternative datum. This is based using the average χ^2 confidence level across all offsets. Table 6.13 lists the average confidence limit for each test piece. For slice only look-up the ROC threshold for TFM1 has a relative magnitude (R_m) of twice that of the average confidence level across all offsets of the test piece.

Table 6.14 compares the results of thresholding using the KI, Otsu and ME techniques

145

Test Piece	$\frac{\sum_{i=1}^{offsets} \chi^2(pc_i, 0.95)}{offsets}$	Slice only R_m	Slice and offset R_m
TFM1	47.7	1.99	1.86
TFM2	47.2	5.8	1.16
TFM3	35.7	5.36	1.77

Table 6.13: ROC thresholds as a multiple of average χ^2 confidence level

to the theoretical optimums suggested by ROC analysis. These are now represented as a percentage change when compared to the previously acquired ROC thresholds.

From table 6.14 the KI thresholds produce results (in terms of accuracy, sensitivity and specificity) closest to those suggested by the ROC analysis. However in most cases the indication is that the KI threshold is high. This is confirmed by a comparison of the actual additional thresholds expressed as a percentage of the dynamic range and listed under the $\Delta th_{\%dR}$ column of the table. For completeness images of the results are illustrated at the end of this chapter in figures 6.34, 6.35 and 6.36.

Test Piece	Method	R_m	Accuracy (%)	Sensitivity	Specificity	F1
TFM1	D^2	1	92.0	1.0	0.92	0.07
	ROC	1.86	98.4	0.99	0.98	0.29
	KI	1.44	97.7	0.99	0.98	0.21
	Otsu	54.97	99.8	0.34	1.0	0.48
	ME	18.03	99.8	0.61	0.99	0.64
TFM2	D^2	1	92.0	0.94	0.92	0.03
	ROC	1.16	95.3	0.94	0.95	0.05
	KI	1.39	97.0	0.88	0.97	0.08
	Otsu	104.79	99.9	0.31	1.0	0.42
	ME	15.86	99.8	0.56	0.99	0.44
TFM3	D^2	1	85.5	0.94	0.85	0.12
	ROC	1.77	94.5	0.92	0.95	0.27
	KI	1.6	93.7	0.92	0.94	0.24
	Otsu	99.1	99.2	0.27	0.99	0.41
	ME	78.0	99.2	0.35	0.99	0.5

Table 6.14: Classification statistics for different thresholds

6.31.3   Example projections after thresholding

A similar study with sectorial data, which did not contain a front wall, demonstrated that image creation through PCA projections significantly enhanced the SNR of each anomaly. In this case it is not possible to use this orientation to measure an anomaly's SNR prior to PCA, therefore this comparison is not possible. It is, however, possible to report each

anomaly's SNR after projection.

Firstly figure 6.33 indicates the areas used for the SNR measurements. Results are presented in tables 6.15 (based on peak foreground value) and 6.16 (based on average foreground value).

Test Piece	F1	F2	F3	F4	F5
TFM1	22.9	35.1	56.27	-	-
TFM2	46.3	53.3	57.1	-	-
TFM3	51.6	27.0	30.8	24.1	14.5

Table 6.15: Peak SNRs (dB) for each fault using manual data sets

a) TFM1   b) TFM2

c) TFM3
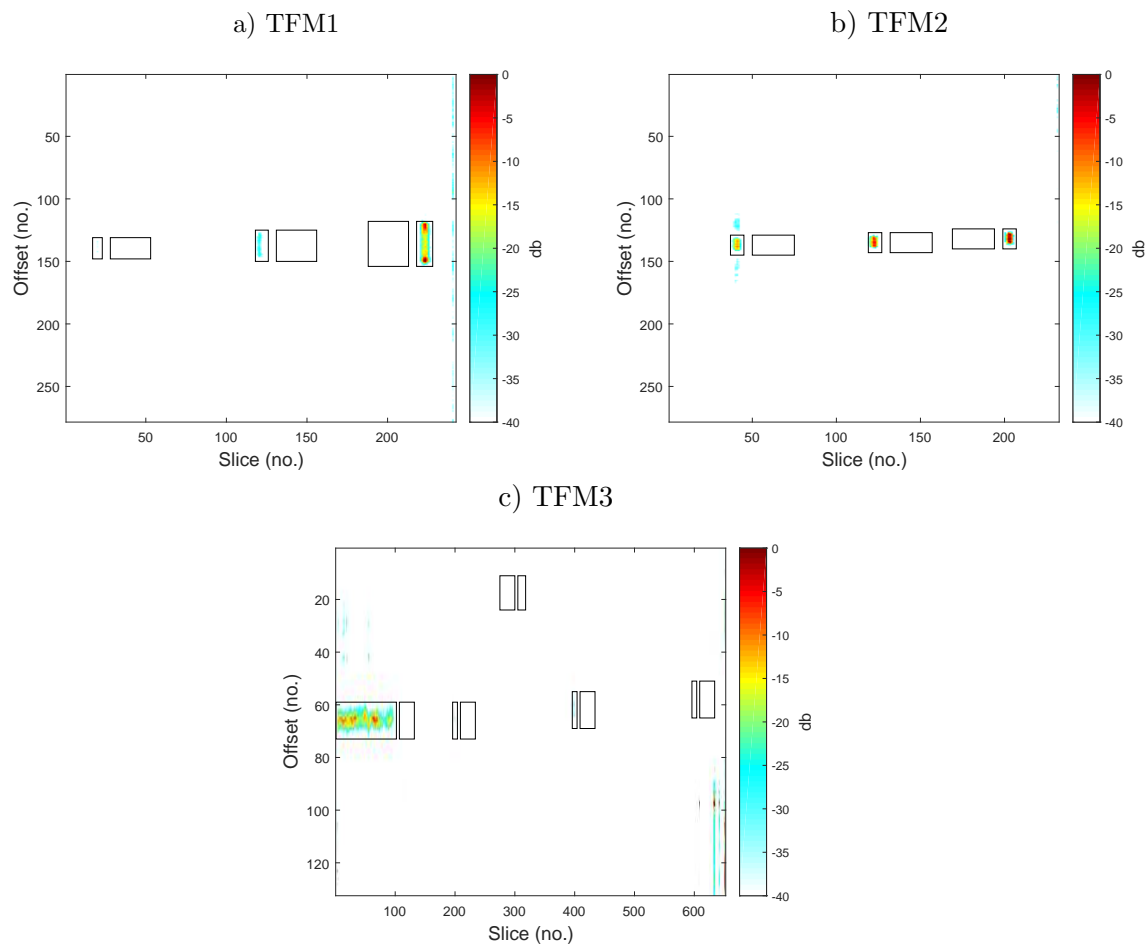
Figure 6.33: Areas for SNR measurements - TFM immersion data sets

Test Piece	F1	F2	F3	F4	F5
TFM1	12.1	23.4	40.6	-	-
TFM2	36.2	42.1	45.7	-	-
TFM3	51.6	27.0	22.7	12.4	5.7

Table 6.16: Average SNRs (dB) for each fault using manual data sets

6.32 Chapter summary

This chapter has demonstrated the potential of PCA for the detection of anomalous conditions in welds from ultrasonic data. All examples are from data collected by phased arrays using either full sector scans or full matrix capture. For sectorial data a sector image is the collection of a set of equal length A-scans each from a single angle and collectively sweeping the sector. In the case of full matrix capture, images are formed using TFM algorithm.

With sectorial scans at oblique incidence, as in the case of TB1 and TB2, major reflections from the first front wall are far removed from the region of interest and can be gated out. In the case of TFM images the front wall is a dominating feature that presents a major obstacle to anomaly detection when taking the preferred plan view. This problem is further exacerbated when a weld cap is present. Attempts at removing the front wall from the TFM data set are not satisfactory. In particular too many assumptions have to be made making resulting software cumbersome, obfuscated and unreliable. For this situation PCA has shown considerable promise. In particular, constant offset projections using standardised data appear to be resilient to the presence of front and back walls as well as a weld cap. This resilience also applies to gradual variations in these features from image to image. The conclusion is that this orientation with standardised data provides the most satisfactory method of anomaly detection. Other orientations for TFM images will be given no further consideration. Full image orientation simply does not provide satisfactory results. Within the main body of the material the constant depth orientation is not affected by the front or back wall. However as the depth approaches one or the other the orientation becomes less reliable; the location of either wall is not constant, there may be skew and the approach fails completely in the presence of a weld cap.

For sector scans the results (table 6.3) indicate that greater accuracy is achieved using constant A-scan rather than full sector projections. The full sector approach has the further disadvantage of producing high dimensional data sets (A-scan samples $\times$ A-scans per sector). However, unlike the TFM case where full image projection show little promise, the results of full sector image projections suggest that some further investigation should be considered.

In the case of TFM and sectorial scans high accuracy is required for the technique to have industrial use. It is demonstrated that accuracy alone is not a satisfactory measure of performance. The results recorded here indicate many instances of high accuracy. However this is not always accompanied by a high probability of anomaly detection, as expressed by sensitivity. In other instances high accuracy is also accompanied by a low specificity indicating that there are too many false positives for the technique to be useable.

Although PCA is demonstrated to have some potential in this context the training sets are, in all cases, selected manually. With careful selection there is some guarantee that these are free from major outliers. Under these conditions the resulting evaluations are likely to demonstrate a good level of discrimination. Not only, therefore, have these results demonstrated the potential of PCA for anomaly detection in this context but they have also provided a level of performance with which following trials can be compared.

For PCA to be useful in the context of anomaly detection it must be possible to obtain the training set automatically from the full data set. In essence this is the subject of the next two chapters. The first approach, discussed in the next chapter, uses data trimming whereby PCA is first used to identify major outliers in the original data set. Ejecting these from the original data set leaves behind observations that have a greater probability of being anomaly free. Repeating PCA using the training set produces a lower rank and more accurate representation of the anomaly free background. A second projection of the full data set against this low rank subspace will now highlight the previously ejected outliers but, more importantly, reveal other outliers masked in the original projections. Results from data trimming may now be compared with those obtained here.
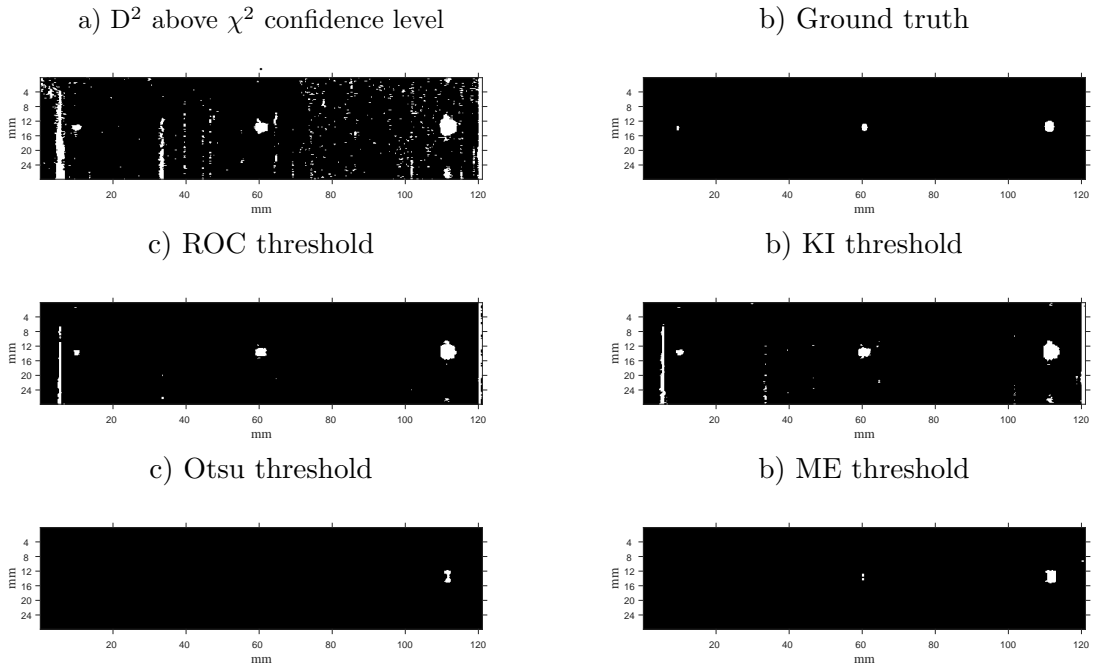
a) D^2 above χ^2 confidence level

b) Ground truth

c) ROC threshold

b) KI threshold

c) Otsu threshold

b) ME threshold

Figure 6.34: TFM1 D^2 above χ^2 confidence level with additional thresholding

a) D^2 above χ^2 confidence level

b) Ground truth

c) ROC threshold

b) KI threshold

c) Otsu threshold

b) ME threshold

Figure 6.35: TFM2 D^2 above χ^2 confidence level with additional thresholding

150

a) D^2 above χ^2 confidence level

b) Ground truth

c) ROC threshold

b) KI threshold

c) Otsu threshold
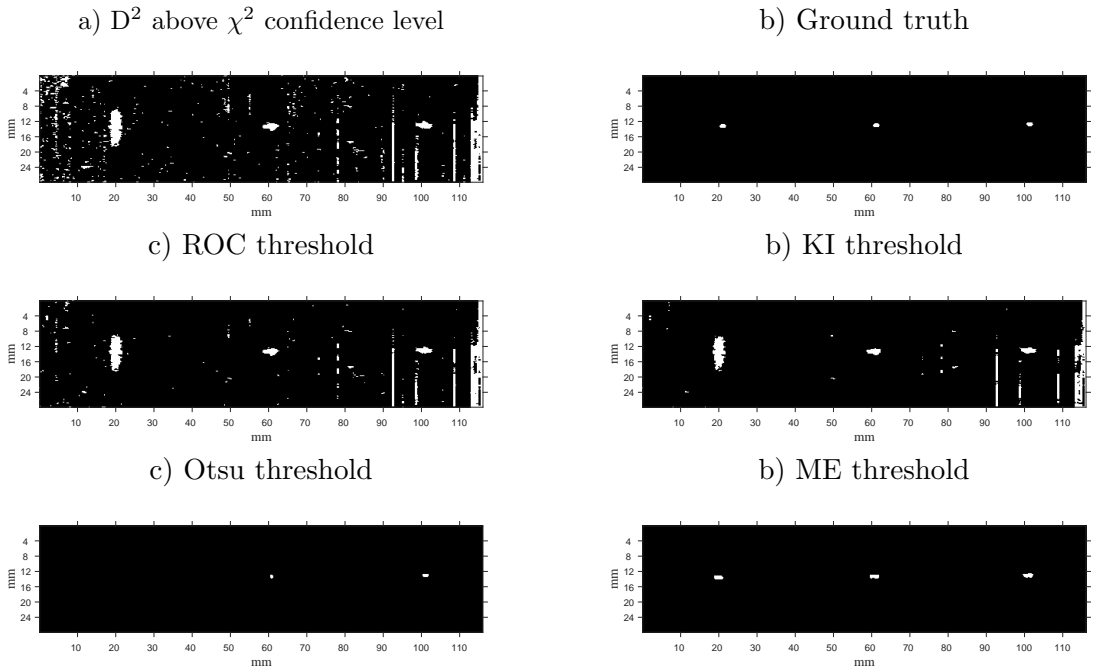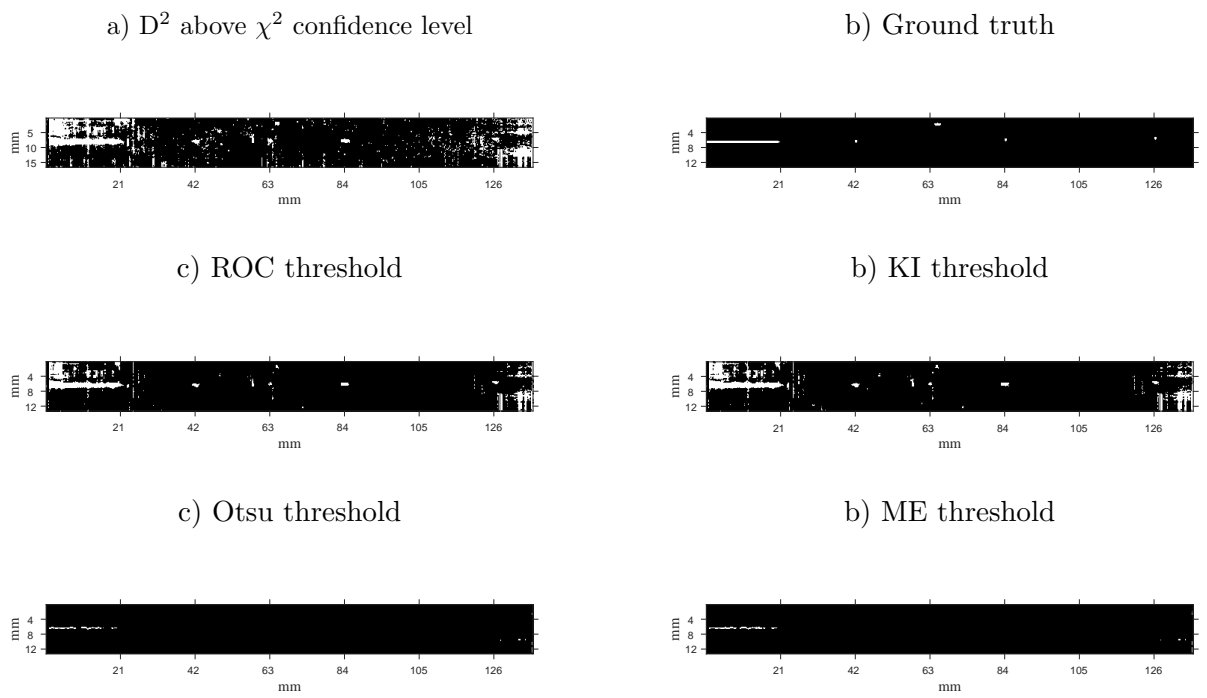
b) ME threshold

Figure 6.36: TFM3 D^2 above χ^2 confidence level with additional thresholding

Chapter 7

Data set trimming

In the case of large regular shaped geometries it is frequently possible to automate the acquisition of ultrasonic inspection data. This is usually achieved by incrementally indexing a probe along a line parallel to the weld. At each index a full scan is taken for off-line processing and inspection by a qualified inspector. Examples of full scan data are the sector and TFM images presented earlier. A significant drawback of this approach is the amount of data produced, all of which must be manually inspected and much of which is expected to be fault free. It has now been demonstrated that PCA is a potential method of automatically highlighting regions of a weld that have some statistical anomaly. However this first requires many examples of normal, anomaly free, observations. Previously these have been selected manually using prior knowledge of the test piece. In practice, a method of obtaining these examples automatically and with no prior knowledge of fault location is essential. This chapter examines methods of doing this.

7.1   Overview

The simplest way of obtaining a training set automatically is to use the full set of observations. This may, of course, result in some examples containing outliers that do not represent the anomaly free distribution. If, however, most of the data set contains normal background observations there is some possibility that they will outweigh the minority of anomalous examples. The proportion of contamination in a data set before outliers can be identified is known as the breakdown point.

Unfortunately, as an unsupervised method of detecting multivariate outliers, the perfor-

mance of PCA is known to degrade in the presence of outliers [96] leading to a low breakdown point with a high number of false indications. Data set contamination manifests itself in a number of ways. Some examples include one, or more, large outlying observations that may skew the mean and covariance towards them. Although these may be identified as outliers they may also mask smaller outliers that are subsequently unidentified. Alternatively a number of outlying instances may collectively skew the mean and covariance estimates, this time making anomalous free observations look like outliers. These two conditions are referred to respectively as masking and swamping.

7.2   Training set anomalies

Before investigating more robust approaches a simple demonstration highlights the effect of outliers on PCA. The demonstration also illustrates how outliers, in higher dimensional data sets, may be identified and subsequently removed.

A scatter plot of the mean centred values of the two variables listed in table 7.1 reveal a

X1	=	-6	-5	-4	-3	-2	0	1	2	3	4
X2	=	-5	-4	-3	-2	-1	1	2	3	4	5

Table 7.1: Example 2 dimensional data without outliers

straight line; figure 7.1.a. As expected the data is represented by a single eigenvector e_1 (0.707, 0.707 or 45°). The two dimensional data can be projected and recovered without error using a single principal component. The same exercise is now repeated with an outlier added to the original data set, table 7.2.

X1	=	-6	-5	-4	-3	-2	**10**	0	1	2	3	4
X2	=	-5	-4	-3	-2	-1	**0**	1	2	3	4	5

Table 7.2: Previous 2 dimensional data with added outlier

The result is a biasing of the first principal component and the mean centring of the plot, figure 7.1.b. Any projections against the first principal component are now corrupted by the presence of the outlier.

Identifying and removing any anomalous observation is one method of circumventing this problem. Firstly, it is noted that, in the case of this extreme outlier, the scatter plot itself provides evidence of an outlier and in practice, inspection of the raw data itself may be

a) Outlier free samples

b) Samples with outlier

c) Reconstruction from PC 1

d) Outlier Identification
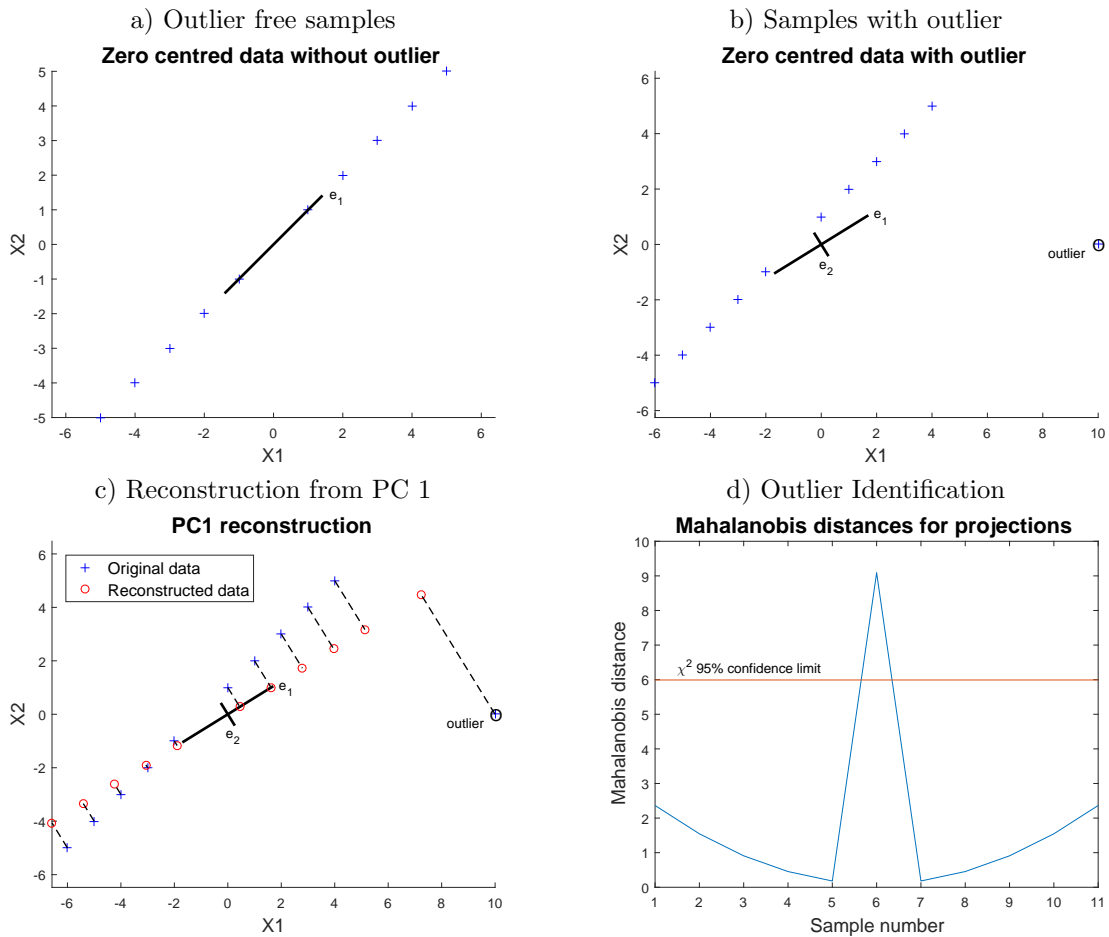
Figure 7.1: Biasing effect of an outlier

sufficient for detection. Where outliers are not so obvious principal components can be examined. For example, Gnanadesikan and Kettenring [97] recommend graphical methods using bivariate plots of pairs of principal components. Gnanadesikan [69] suggests that the last few, as well as the first few principal components are usually of interest, although the first few principal components are especially sensitive to outliers as they may inappropriately inflate variances and co-variances. However, as the number of dimensions increase, these manual methods are not so straightforward and, more importantly, automation requires an analytic approach.

7.3 Robust estimation

For PCA a 'robust fit' refers to the process of finding an estimation of the eigenvectors that would exist without contamination. Outliers are now identified by their distance scores from this fit. In the previous chapter this was attempted by manually selecting a data set free from anomalies. Figure 6.7.a highlights the existence of statistical differences indicating this visual inspection to be unsatisfactory. The limitation is further highlighted by Hubert et al. [98] who discuss alternative approaches to obtaining a robust estimate. Principal amongst these are the Maximum Covariance Determinant (MCD) and Minimum Volume Ellipsoid. The two have similarities; however a principal difference is that for a predefined number of points, h, 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 [99], [100] and Rousseeuw and Leroy [101]. 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 [98], [102]. To give an indication of the procedure the following gives an outline of MVE. That for MCD is very similar and is provided in the previously cited publications.

7.4 MVE

MVE involves using an iterative algorithm which, for a large data set, requires intensive sampling and computation. An ellipsoid that inflates or deflates on each iteration contains exactly h (out of n) points. For example $p+1$ observations are repetitively and randomly drawn from the data set. For each set of observations, indexed by J, the median squared, m_J^2, where med is the median for $i = 1$ to n and $h = (n+p+1)/2$ is determined:-

$$m_J^2 = med[(x_i - \bar{x}_J)C_J^{-1}(x_i - \bar{x}_J)^T]h : n \tag{7.1}$$

For each J a value P_J, which is proportional to the resulting ellipsoid is calculated:-

$$P_J = \sqrt{det(m_J^2 C_J))} \tag{7.2}$$

After a number of repetitions two estimated MVE terms representing the mean:-

$$T(X) = \bar{x}_J$$

155

and the covariance:-

$$C(X) = \frac{c^2(n,p)}{\chi^2_{p,0.5}} m_J^2 C_J$$

are calculated for the set with the lowest P_J.

(Note: $c^2(n,p)$ is a small sample correction term given as $[1 + 15/(n\text{-}p)]^2)$ and $\chi^2_{p,0.5}$ is the median of the $\chi^2_{(pc,0.95)}$ distribution with p degrees of freedom)

A statistic, similar to the Mahalanobis distance, can now be calculated for observation i:-

$$W_i = (x_i - T(X))C(X)^{-1}(x_i - T(X))^T \tag{7.3}$$

If $W_i > \chi^2_{p,0.975}$ the observation is defined as an outlier. PCA is now done with the outlying identifiers removed or assigned a weight of 0.

7.5 Mahalanobis vs MVE

Using simulated data Jackson and Chen [103] were able to compare the performance of the Mahalanobis and MVE metrics in terms of their ability to detect outliers. With a small number (up to 10%) the Mahalanobis metric is reported to perform better than MVE. In fact, under conditions of a few or no outliers, MVE identified the most observations as being outliers, but which were actually false positives. As the percentage of outliers increases to around 30% both approaches become similar. With further contamination the Mahalanobis measure gradually deteriorates whilst the MVE improves, eventually becoming the superior.

Similar results are reported by Kosiski [104] and Marden [105]. Clearly no approach is satisfactory if the data set contains 50% or more contamination. An advantage of MVE is that it performs better than the Mahalanobis measure as the level of data contamination approaches the 50% limit.

In contrast to the sectorial data which may contain over 1500 variables the data sets used in Jackson and Chen's investigations involved simulated data with only 4 variables and 50 observations. The validity of the results when applied to large scale real data is not questioned; however data set size now makes the MVE approach less attractive in terms of computation time. This is particularly the case if data from each weld is treated individually. MVE is a preferred option if the background statistics are derived from a single inspection piece and used repeatedly on data sets from similar welds. Where many

welds with slightly different geometries, material or test equipment are to be used, it is more likely that unsupervised anomaly detection using the individual data set will lead to more accurate results. In this case the Mahalanobis metric, which is computationally far more efficient, has advantages.

Under the limitations of time and available data this investigation will initially use the Mahalanobis distance. This measure has found value in other applications [82] and is, computationally, relatively undemanding. Other than a confidence level Mahalanobis distance requires no further parameters. Although Rousseeuw [106] provides a formula to determine h (used in equation 7.1), previous research [103] indicates some dependence of MVE's accuracy on this value which has to be increased.

A more recent alternative to obtaining an accurate estimation is Robust PCA (RPCA) proposed by Candès et al. [107]. Like the MVE and MCD approaches this suffers from high computational costs. However it represents a significant advance. In the case of ultrasonic imaging or similar applications no evidence of its use has yet been found. A discussion of the approach, its potential application and presentation of results, is left to later chapters in this work.

7.6   Robust estimation using the Mahalanobis metric

Examples of robust PCA using Mahalanobis, MVE and MCD are found in many applications relating to on-line fraud and intruder detection systems [82], [108] and [109]. Of these Shyu et al. [82] use multivariate trimming to obtain the robust co-variance estimator. The approach is, firstly, to use a Mahalanobis metric to identify a percentage of extreme observations to be trimmed (removed) from the training set. Under the assumption of a $\chi^2_{(pc,0.95)}$ distribution, and using equation 6.15, an observation x is an outlier if, for the first q principal components:-

$$\sum_{i=1}^{q} \frac{y_i^2}{\lambda_i} > \chi_q^2(\alpha) \qquad (7.4)$$

New, trimmed, estimators of both the mean and covariance matrix ($\bar{x}$ and Σ) are then calculated from the remaining observations. After trimming, some suspicious data may remain; however the assumption is that the number of background instances will be much larger than the number of anomalies. If this is the case anomalous observations become more distinct from the central cluster of background observations. Trimming can be repeated more than once but in the cited work repeated trimming was limited to maintain

the number of observations greater than the number of dimensions (ie $n > p$). In this case the sectorial data is HDSLL, limiting the maximum number of PCs to the number of observations so that repeated trimming is not recommended.

The authors also select a value for the significance level, α, of 0.95 and this is the value used here unless otherwise specified.

7.7 Sectorial data set selection

Four PA runs (PA1 to PA4) for each test block (TB1 and TB2) provide a total of 8 individual data sets. All are captured using the same phased array, instrumentation and settings. The most likely differences between each test run are now due to manual manipulation. Although precautions are taken to ensure the same physical operation slight variations between each might include, for example, offset (distance between probe and weld centre line) and skew (slight rotation of the phased array as it is guided along the test piece).

In total 1816 sectors (227 sectors from each PA direction) are recorded. It is known that some of these, particularly at the beginning and end of the data acquisitions are invalid. A more careful examination of the data suggests a total of approximately 1630 valid data sets. This is more than the theoretical maximum of 1600 (section 4.4) and is thought to be due to the presence of duplicate entries. As identification of these is difficult it was decided, for this exercise, not to interfere with the data set, the original 1816 sectors being maintained. A key issue is the fact that, in comparison to the previous selection of 700 observations, there is now no attempt to manually interfere with the data set.

The ground truth information indicates that approximately 500 sectors contain anomalies so that in total approximately 1130 sectors (1630 - 500) or 40% of the full data set are anomalous in one way or another. This is above the upper end of the limit for the Mahalanobis metric suggested by Jackson and Chen [103]. However after trimming, the anticipation is that although the data size of the overall training set will reduce so too will the percentage of outliers. That is, the observations ejected will be those containing some form of anomaly.

Although, in this case, combining data from all 8 (PA) passes of the two test blocks produces the largest test set, in practice, situations may occur where only a single run of a weld exists. In other cases more runs may be available but variations in set up, due to physical constraints and accessibility, may be so great that their combination does not provide a representative set of samples. For this reason it is necessary to consider

the individual passes, each containing 227 sectorial scans, as well as the larger data set containing the combination of all 8 passes. Although the individual passes will, on average, contain the same density of anomalies as the combined data set their smaller size is likely to reduce classification accuracy. Nonetheless the comparison is made and conclusions drawn.

Before any comparison between the combination and individual data sets is made consideration is first given to the potential effects of trimming. This example uses the full test set of 1816 sectors. In addition to providing results for comparison it provides further details of the procedure.

7.8 Trimming - full sector projections

Figure 7.2.a and b illustrate the distribution of the first 3 PC scores of the entire untrimmed data set and the individual Mahalanobis distance (D^2) of each sector. The larger Mahalanobis distances represent anomalous outliers. After sorting into ascending order, figure 7.2.c, sectors with the largest Mahalanobis distances are removed. In this case the cut-off point is the $\chi^2_{(pc,0.95)}$ confidence level. Once again this level is chosen for consistency with other investigations. The number of PCs to use is determined from the Kaiser stopping rule. Further details of this follow in section 7.8.1. The remaining 1477 observations (figure 7.2.c) form a new training set where figure 7.2.d illustrates the new distributions of the first 3 PC scores. This new, tighter, cluster results from the omission of the outliers but more important is an improvement in the accuracy of the mean values and the co-variance matrix. The net effect is that projection of a unknown observation will result in a more accurate discrimination between anomalous and anomaly free conditions.

To illustrate the effect of this on anomaly detection, figure 7.2.e and f contains a plot for the Mahalanobis distances of each sector from test piece TB1PA1, when projected using the eigenvector coefficients for the untrimmed and trimmed training sets respectively.

Initially the confidence limits for both trimming and detection are set at $\chi^2_{(pc,0.95)}$. Under these conditions the figures indicate that with the trimmed training set all anomalies are detected at least once. In the case of the untrimmed training set two anomalies are masked out and are consequently undetected.

In addition to giving an indication of all six intended anomalies, trimming also enhances the Mahalanobis distance measure, further improving the probability of detection. As usual this is at the expense of an increase in the number of false positives. If this increase results from a widening of the base of each true peak, to include sectors just outside the edge of

a) Projections before trimming

b) D^2 before trimming

c) D^2 in ascending order

d) Projections after trimming

e) D^2 for TB1PA1 before trimming

f) D^2 for TB1PA1 after trimming
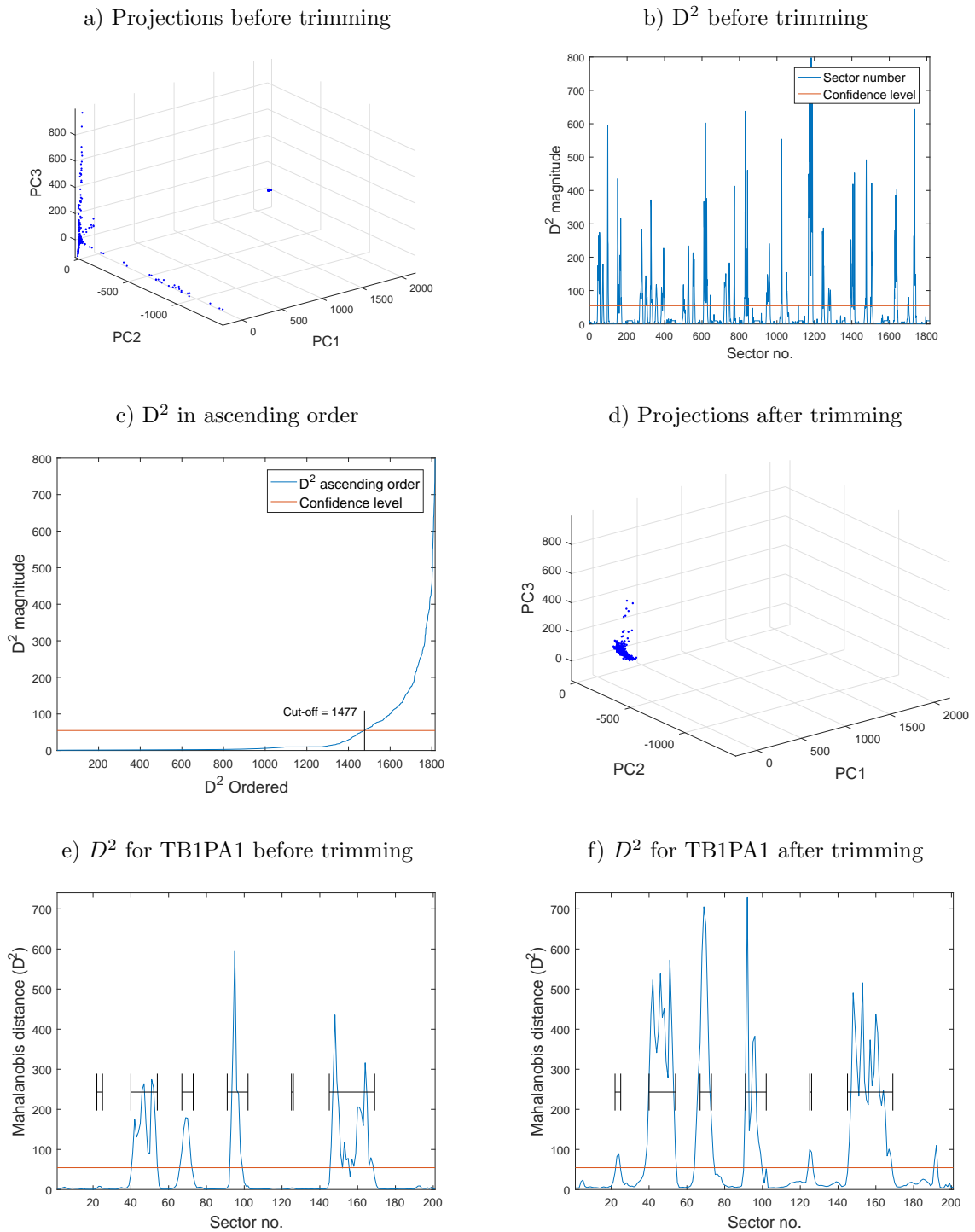
Figure 7.2: Example of data set trimming

the test vector, this is not too serious an issue.

7.8.1   Number of principal components

Before continuing it is necessary to comment on the number of principal components used in each case. Trimming of the original data set uses 39 principal components as suggested by the Kaiser stopping rule. After trimming the rule gave a value of 5 producing accuracy and sensitivity figures far lower than those for the untrimmed example. On examination the two scree plots have a significant 'spike' with a large emphasis on PC1 ($>5.10^5$). For the trimmed data set this drops off far more steeply than that for the untrimmed case. The reason for both shapes is thought to be due to the high dimensionality of each data set. To circumvent the problem it was decided to use the same number of principal components for the trimmed data set; that is, 39.

7.8.2   Untrimmed and trimmed data sets

In all cases projections using the trimmed data indicate an improvement in accuracy over those for the untrimmed case, table 7.3. Not directly evident is the fact that projections using the untrimmed data fail to detect 18 anomalous instances, some of which represent comparatively large and significant planar features such as cracks. In the case of the trimmed data set this reduces to 5 with a stronger indication of each anomaly. In all cases the 5 missed anomalies are smaller pores. Once again this is evident in figure 7.2.

Although accuracy gives some measure of performance it gives no indication of the number of false negatives (failure to detect an anomaly). In this context measures of sensitivity, which indicate the ability to detect true positives, and specificity, which indicate the ability to detect true negatives, provide further information. Table 7.3 lists these figures for the untrimmed and trimmed examples. A further comparison is provided by the F1 score which represents the harmonic mean of precision and sensitivity.

The untrimmed data indicates high figures for specificity (1 indicting no false positives). However the respective sensitivity figures indicate a low number of true positives. For the trimmed data there is a relatively large increase in the sensitivity (overall nearly 30%) indicating a greater detection of anomalies. Although the accuracy figures only slightly favour the trimmed data set the trimmed sensitivity values are far more emphatic. The slight downside is a reduction in specificity but this never falls below 90%.

Test piece	Untrimmed (39 PCs)				Trimmed (39 PCs)			
	Acc.(%)	Sens.	Spec.	F1	Acc.(%)	Sens.	Spec.	F1
TB1PA1	91.0	0.74	0.99	0.84	92.5	0.89	0.94	0.86
TB1PA2	88.0	0.71	0.97	0.83	90.5	0.91	0.91	0.87
TB1PA3	88.5	0.65	1.00	0.82	90.5	0.89	0.91	0.85
TB1PA4	89.3	0.66	1.00	0.81	92.4	0.91	0.92	0.87
TB2PA1	87.3	0.57	1.00	0.80	94.1	0.95	0.94	0.88
TB2PA2	87.8	0.64	1.00	0.80	91.7	0.93	0.91	0.88
TB2PA3	83.4	0.50	1.00	0.78	93.7	0.85	0.97	0.88
TB2PA4	88.8	0.58	1.00	0.78	93.2	0.93	0.93	0.88
Averages	88.0	0.63	0.99	0.81	92.3	0.90	0.92	0.87

Table 7.3: Discrimination before and after trimming

7.8.3 Manual and trimmed data sets

The previous comparison simply confirms that the removal of outliers produces a more accurate estimate of the population's anomaly free mean values and co-variance matrix, in turn leading to improvements in discrimination. Attention now turns to an assessment of trimming as a means of training set selection. In the previous chapter outlier free observations were selected manually. Whilst the manual training set facilitated an initial investigation of the approach, in practice automation of the process is essential. Results of a comparison between the previous manual training set of 700 observations and that due to automatic trimming are now presented. It is first noted that, as indicated in figure 7.2.c, and using the $\chi^2_{(pc,0.95)}$ confidence level, 1477 observations are retained for the trimmed training set. This is over twice that for the manual selection potentially giving the trimmed training set some advantage. However there is confidence that the manually selected training set contains only background observations.

For convenience table 7.4 lists previously presented discrimination results for the manual training set (table 6.1) and the trimmed training set (table 7.3). With the exception of a small reduction in specificity all statistics for the trimmed training set exceed those of the manual training set. At the same time the number of PCs used after trimming is almost half of that for the manually selected set. On this basis not only does trimming provide automatic training set selection, it is an improvement, both in terms of discrimination and computation time, over the manual selection.

162

Test Piece	Manual $\chi^2_{(pc,0.95)}$ (82 PCs)				Trimmed $\chi^2_{(pc,0.95)}$ (39 PCs)			
	Acc.(%)	Sens.	Spec.	F1	Acc.(%)	Sens.	Spec.	F1
TB1PA1	89.1	0.66	0.99	0.80	92.5	0.89	0.94	0.89
TB1PA2	82.7	0.61	0.98	0.74	90.5	0.91	0.91	0.86
TB1PA3	86.6	0.59	0.99	0.75	90.5	0.89	0.91	0.85
TB1PA4	89.3	0.66	0.98	0.79	92.4	0.91	0.92	0.88
TB2PA1	88.2	0.62	0.99	0.76	94.1	0.95	0.94	0.91
TB2PA2	88.2	0.68	0.99	0.78	91.7	0.93	0.91	0.88
TB2PA3	82.4	0.47	0.99	0.64	93.6	0.85	0.97	0.90
TB2PA4	93.2	0.76	0.99	0.86	93.2	0.93	0.91	0.88
Averages	87.5	0.63	0.99	0.76	93.6	0.91	0.93	0.89

Table 7.4: Discrimination comparisons: of manual and trimmed training sets

7.8.4 ROC analysis

To complete this evaluation a further comparison between the two training sets uses ROC analysis, table 7.5. Of particular interest is the fact that at some threshold level there is little difference between the results for each test block. Moreover the overall averages also indicate little difference. This suggests that the trimming procedure is capable of producing a training set with the discrimination ability of one from a careful manual selection. Similarities between the two also suggests some equivalence between the two subspaces whilst the fact that the trimmed training set uses only half the number of PCs suggests some superiority of the estimation.

A further noteworthy point is that for the trimmed training set there is little difference between the discrimination statistics using the original $\chi^2_{(pc,0.95)}$ confidence level and those from its ROC analysis. For the manual selection the difference is wider. Table 7.6 and the example projections for TB1PA1 further emphasise the closeness of the thresholds for the trimmed training set. The implication of this is that, for this approach, little improvement is possible over the $\chi^2_{(pc,0.95)}$ confidence level alone.

7.8.5 Individual test sets

In the discussion, so far, the data set combines all 8 of the original test runs. This has the advantage of maximising the number of available observations. An alternative, more appropriate to those practical conditions where a prior training set is not available, is to use data from a single scan of an individual weld. In practice a weld may often be a few metres in length making many observations available. By contrast each individual data set

163

Test Piece	Manual					Trimmed				
	AUC	Acc.(%)	Sens.	Spec.	F1	AUC	Acc.(%)	Sens.	Spec.	F1
TB1PA1	0.99	93.5	0.98	0.91	0.86	0.98	93.0	0.98	0.90	0.91
TB1PA2	0.95	90.9	0.87	0.93	0.82	0.96	90.4	0.89	0.91	0.88
TB1PA3	0.99	95.2	0.99	0.94	0.85	0.97	91.4	0.84	0.95	0.87
TB1PA4	0.99	94.9	0.93	0.96	0.86	0.98	93.9	0.92	0.95	0.90
TB2PA1	0.98	92.7	0.97	0.91	0.83	0.99	95.6	0.95	0.96	0.93
TB2PA2	0.97	94.1	0.93	0.95	0.87	0.97	95.1	0.91	0.97	0.93
TB2PA3	0.95	92.2	0.93	0.92	0.78	0.98	95.6	0.93	0.97	0.93
TB2PA4	0.99	97.6	0.98	0.97	0.91	0.98	93.7	0.96	0.92	0.88
Averages	0.98	94.0	0.95	0.94	0.85	0.97	93.5	0.92	0.94	0.90

Table 7.5: ROC comparisons: manual and trimmed training sets

Test Piece	Thresholds % of dyn. Range	
	$\chi^2_{(pc,0.95)}$	ROC_{th}
TB1PA1	7.48	5.07
TB1PA2	3.7	5.29
TB1PA3	2.89	1.43
TB1PA4	2.89	1.43
TB2PA1	4.3	5.6
TB2PA2	0.85	1.15
TB2PA3	3.24	1.49
TB2PA4	5.23	3.93
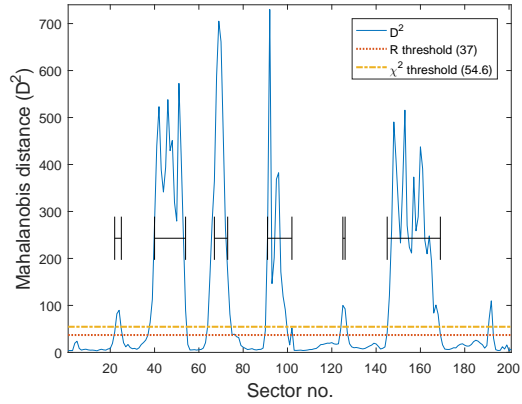
Table 7.6: Threshold comparisons as % of total dynamic range

available here covers a distance of less than 40 centimetres producing a maximum of around 200 observations. These are likely to be insufficient to provide an accurate statistical model of the fault free background, particularly when the faults are densely packed. However within these limitations, and in view of its potential practical application, this approach is given some attention.

Figure 7.3 illustrates the effect, on data set size, of trimming the full and an individual (TB1PA1) data set. The full data set contains 1816 observations. After trimming, using the $\chi^2_{(pc,0.95)}$ confidence level 81% of the original observations remain. This is slightly higher than the 73% for the accompanying individual data set. Investigations for the remaining 7 test pieces produce a similar pattern. In particular they reveal that an average of 72% observations $\pm 3\%$ remain after trimming and that the number of major PCs is 23 ± 5.

It is already established that trimming provides some improvement in anomaly detection. In particular removal of major outliers reduces effects due to masking and swamping.

a) Full training set b) Individual training set (TB1PA1)
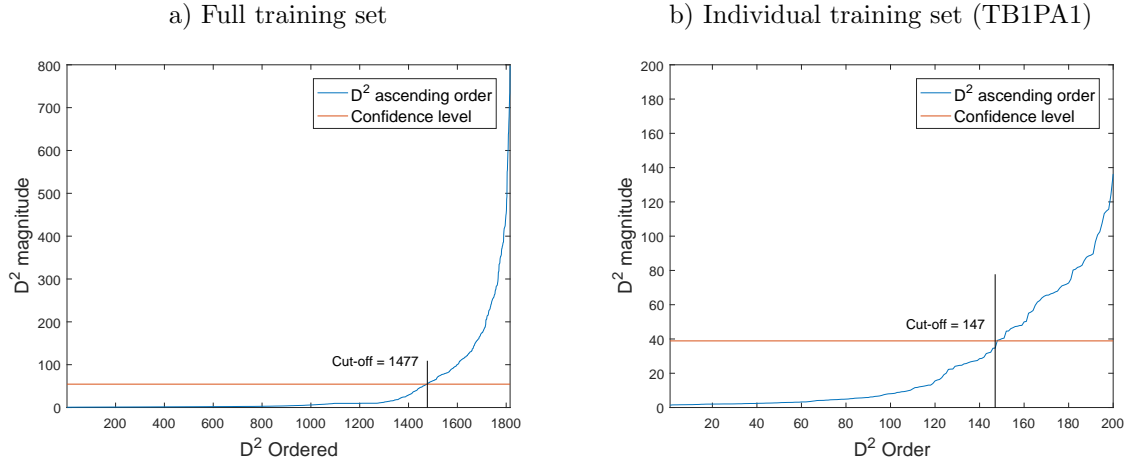
Figure 7.3: Comparison of trimming on full and individual data sets

The first investigation is simply to confirm this for the smaller data sets. The results are presented in table 7.7. The respective $\chi^2_{(pc,0.95)}$ accuracies before and after trimming do not indicate significant differences. However there is, in all cases, an improvement in sensitivity, indicating a higher hit rate after trimming. Although a corresponding reduction in specificity is expected this is smaller than the percentage increase in sensitivity. This is in agreement with the finding for the full data set and provides further evidence that the untrimmed data set tends to mask anomalies.

Test Piece	Untrimmed $\chi^2_{(pc,0.95)}$				Trimmed $\chi^2_{(pc,0.95)}$			
	Acc.	Sens.	Spec.	F1	Acc.	Sens.	Spec.	F1
TB1PA1	90.1	0.77	0.97	0.84	86.5	0.92	0.84	0.82
TB1PA2	81.7	0.61	0.96	0.73	91.8	0.89	0.93	0.90
TB1PA3	89.0	0.71	0.98	0.81	89.5	0.97	0.86	0.86
TB1PA4	90.8	0.74	0.99	0.83	87.2	0.89	0.87	0.81
TB2PA1	89.2	0.67	0.99	0.79	93.6	0.98	0.92	0.90
TB2PA2	87.8	0.64	1.00	0.78	94.1	0.94	0.94	0.92
TB2PA3	89.3	0.72	0.98	0.82	87.8	0.88	0.88	0.83
TB2PA4	92.7	0.82	0.97	0.86	82.5	0.85	0.81	0.81
Averages	88.8	0.71	0.98	0.81	89.1	0.92	0.88	0.85

Table 7.7: Discrimination before and after trimming, individual training sets

To further this investigation the last 5 columns of table 7.8 list the results of a comparison between the trimmed results and those from ROC analysis. The results have many similarities to those for the full data set (table 7.5). In particular the ROC AUCs suggest good classification whilst the F1 statistics provide confidence in the analysis. Once again the potential improvements indicated by the ROC analysis represent only slight improvements

165

Test Piece	Trimmed $\chi^2_{(pc,0.95)}$				ROC				
	Acc.	Sens.	Spec.	F1	AUC	Acc.	Sens.	Spec	F1
TB1PA1	86.5	0.92	0.84	0.82	0.96	92.5	0.92	0.93	0.89
TB1PA2	91.8	0.89	0.93	0.90	0.97	91.3	0.88	0.94	0.90
TB1PA3	89.5	0.97	0.86	0.86	0.96	92.3	0.96	0.91	0.89
TB1PA4	87.2	0.89	0.87	0.81	0.85	87.2	0.84	0.90	0.82
TB2PA1	93.6	0.98	0.92	0.90	0.98	94.6	0.97	0.94	0.92
TB2PA2	94.1	0.94	0.94	0.92	0.97	94.1	0.96	0.93	0.92
TB2PA3	87.8	0.88	0.88	0.83	0.95	88.3	0.88	0.88	0.84
TB2PA4	82.5	0.85	0.81	0.81	0.91	88.8	0.86	0.90	0.81
Averages	89.1	0.91	0.88	0.85	0.94	91.1	0.91	0.92	0.87

Table 7.8: ROC analysis, individual data sets (trimmed)

leading to the conclusion that significant improvement by further thresholding alone is not possible.

Finally table 7.9 compares the statistics from the two trimmed training sets (individual and full). This more clearly shows the similarities between the two categories, there being some slight advantage in those from the full data set. The similarities are likely to be attributed to the fact that in terms of dimensionality ($1544 \times 71 = 109{,}624$) the number of observations is, in both cases, extremely low ($p \gg n$). The slight advantage of the full data sets is most likely due to its higher proportion of anomaly free sectors after trimming (e.g. 81% vs 73%).

Overall these results indicate that full sector projections provide some potential for anomaly detection. However to improve the probability of detection the training set needs to contain far more anomaly free examples. The data sets available here are both short and contain a high density of anomaly. In practice if a single weld contains large sections of anomaly free observations it is likely that an individual data set will provide a high probability of detection. With the present data sets this cannot be confirmed.

7.9 Constant A-scan trimming

This method is outlined in sections 6.9 and 6.24 where the advantages and drawbacks of the approach are also discussed. The procedure is repeated here, this time after trimming the whole data set rather than making a manual selection. The original intention was to compare the results after trimming with those before trimming. However it was quickly realised that trimming provides a significant improvement over discrimination using the

Test Piece	Full reference data set				Individual reference set (TB1PA1)			
	$\chi^2_{(pc,0.95)}$	sens.	spec.	F1	$\chi^2_{(pc,0.95)}$	sens.	spec.	F1
TB1PA1	92.5	0.89	0.94	0.89	86.5	0.92	0.84	0.82
TB1PA2	90.5	0.91	0.91	0.86	91.8	0.89	0.93	0.90
TB1PA3	90.5	0.89	0.91	0.85	89.5	0.97	0.86	0.86
TB1PA4	92.4	0.91	0.92	0.88	87.2	0.89	0.87	0.81
TB2PA1	94.1	0.95	0.94	0.91	93.6	0.98	0.92	0.90
TB2PA2	91.7	0.93	0.91	0.88	94.1	0.94	0.94	0.92
TB2PA3	93.6	0.85	0.97	0.90	87.8	0.88	0.88	0.83
TB2PA4	93.2	0.93	0.91	0.88	82.5	0.85	0.81	0.81
Averages.	92.3	0.91	0.93	0.88	89.1	0.92	0.88	0.85

Table 7.9: Comparison of full and individual reference sets after trimming

whole untrimmed reference set. In the case of the constant A-scan orientation further evidence of improvement from trimming is given by figure 7.4. Here the results of a sector only look-up and sector with A-scan look-up are compared before and after trimming. Before trimming (figures 7.4.a and b) the Mahalanobis distances are not particularly distinct and many projections remain above the thresholds. Although untrimmed projections can produce high accuracy figures, these rarely exceed 85% and contain many false positives which reduce specificity whilst increasing sensitivity. By comparison the trimmed projections (figures 7.4.c and d) are much sharper giving a more distinct discrimination. Overall, the trimmed reference set produces a more accurate representation of the anomaly free conditions. Consequently, and for the sake of brevity, no further comparisons will take place between the untrimmed and trimmed reference sets.

For constant A-scan trimming two look-up methods are compared. The first compares the confusion statistics based on sector only look-up, the second uses both sector and A-scan look-up. Results of both methods are presented in table 7.10.b. For convenience the results from manual selection of the reference data (table 6.3), are repeated (table 7.10.a). It is with some surprise that the trimmed projections represent some degradation over those obtained by manual selection with those for sector only look-up being particularly significant.

A contributory reason for this may be gleamed from a comparison of figure 7.4.c and 6.12.b. As a reminder both of these curves are plotted after subtracting the threshold so that any value greater than 0 represents a potential anomaly. With trimming, the Mahalanobis distances are approximately $\times 10$ greater than for the manually selected data set. Although this enhances the anomalies it is also accompanied by an increase in disturbances on the floor of the plot, corresponding to areas free from anomalies. The conjecture is that for the

	Look up method					Look up method			
	Sector only		with A-scan			Sector only		with A-scan	
	TB1	TB2	TB1	TB2		TB1	TB2	TB1	TB2
PA1	55.2	52.5	92.4	89.6	PA1	40.8	36.8	83.9	81.3
PA2	45.2	44.8	85.8	89.3	PA2	43.8	37.8	69.5	81.3
PA3	62.7	66.3	91.0	93.3	PA3	46.9	49.8	84.7	89.0
PA4	65.3	59.2	91.7	93.3	PA4	42.4	40.8	86.0	87.4
	(a) Manual Training set (700)					(b) Trimmed training sets			

Table 7.10: % Accuracy manual (a) and trimmed (b) constant A-scan projections

trimmed data set the anomalies are more pronounced, increasing the probability of detection, but the numerous low level projections breaking through the threshold significantly increase the number of false positives, thereby compromising the accuracy. One method of improving this is to increase the $\chi^2_{(pc,0.95)}$ confidence level. For a well-characterised application this may well be a solution. However for this assessment the $\chi^2_{(pc,0.95)}$ value is maintained for comparisons.

7.9.1 Sector only look-up

Accuracy figures for sector only look up are poor and average around 40%. The sector only look up curve, figure 7.4.c, is generated by taking the maximum Mahalanobis distance (across all A-scans) for each sector. This increases the probability of a sector containing a value greater than the threshold. For the particular example (figure 7.4.c) over 50% of sectors contain a false positive. This is evident from the fact that the curve rarely reduces to the 0 baseline. ROC analysis is now employed to investigate the potential of this approach.

7.9.2 Sector only ROC analysis

Table 7.11 summarises the accuracy results for all test pieces. In all cases the AUC remains high indicating the classifier's potential with the F1 scores giving further confidence.

The ROC accuracies represent a significant improvement over the original $\chi^2_{(pc,0.95)}$ values. In all cases they exceed 94.0% with high values for both sensitivity and specificity. A further observation is that all of these theoretical accuracies are greater than those of the original full sector projections reported in table 7.5 and its corresponding ROC analysis,

a) Sector only look-up (untrimmed)

b) A-scan and sectors look-up (untrimmed)

c) Sector only look-up (trimmed)

d) A-scan and sectors look-up (trimmed)
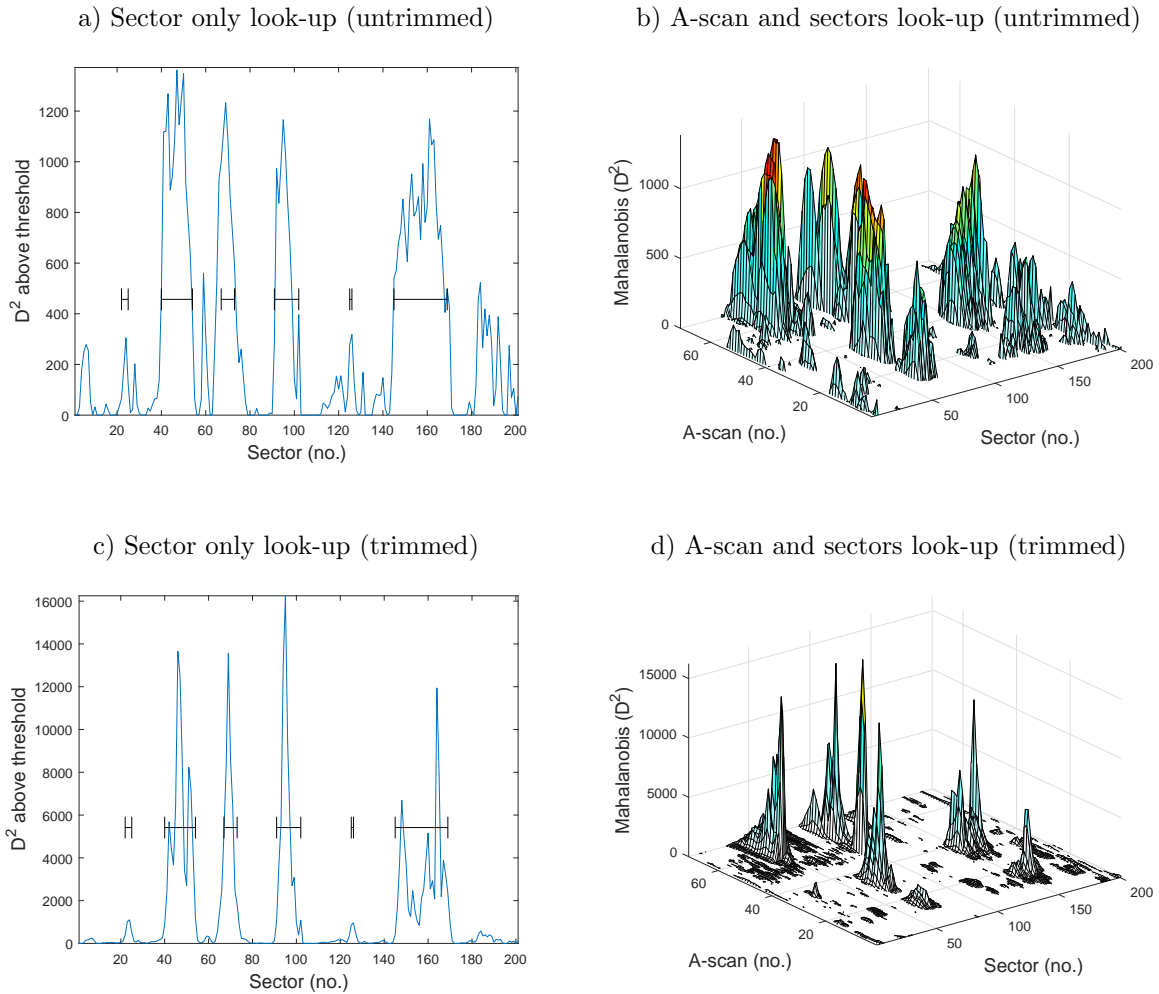
Figure 7.4: Effect of trimming on D^2 projections (constant A-scan)

table 7.4. A similar comparison of sensitivity, specificity and F1 score again demonstrates consistent improvements; however these remain theoretical.

Despite not being able to provide a threshold for automatic anomaly detection accentuation of the Mahalanobis distance does provide the inspector with a visual tool that highlights areas containing potential anomalies. For example, figure 7.4.c contains distinct peaks corresponding to sectors containing each anomaly. Smaller peaks, not associated with an implanted anomaly, indicate sectors where there is some non-intended imperfection or some slight instrumentation disturbance affecting the actual measurements. For example,

169

a) Original projections b) Projections with $\chi^2_{(pc,0.95)}$ subtracted
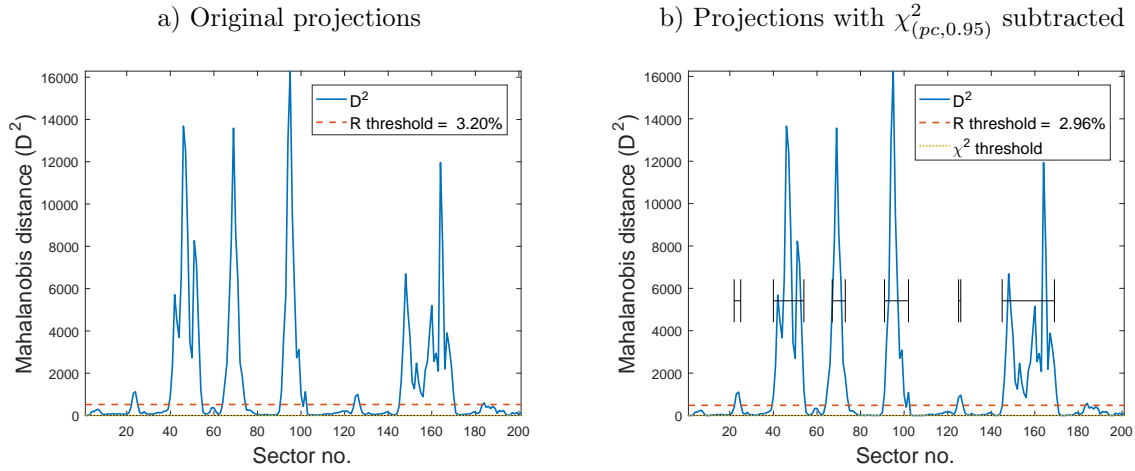
Figure 7.5: ROC threshold with and without $\chi^2_{(pc,0.95)}$ level subtracted (TB1PA1)

the small peaks around sector positions 5 and 190 are frequently present in the sectorial data. They do not correspond to any real anomaly in the material and are thought to be caused by a disturbance to the couplant whilst the probe is momentarily stationary. Alternatively, they are reflections from the side of the test piece due to beam spread in the passive direction.

Test Piece	$\chi^2_{(pc,0.95)}$				ROC				
	Acc.(%)	sens.	spec.	F1	AUC	Acc.(%)	sens.	spec	F1
TB1PA1	40.8	1.0	0.13	0.52	0.99	95.5	0.94	0.96	0.92
TB1PA2	43.8	1.0	0.05	0.59	0.99	96.2	0.94	0.98	0.95
TB1PA3	46.9	1.0	0.21	0.55	0.97	95.2	0.99	0.94	0.93
TB1PA4	42.4	1.0	0.16	0.52	0.99	96.9	0.97	0.97	0.95
TB2PA1	36.8	1.0	0.10	0.49	0.99	96.1	0.98	0.95	0.94
TB2PA2	37.8	1.0	0.07	0.52	0.96	95.1	0.93	0.96	0.93
TB2PA3	49.8	1.0	0.25	0.57	0.99	94.6	1.00	0.92	0.93
TB2PA4	40.8	1.0	0.19	0.47	0.98	94.7	0.93	0.94	0.90
Averages	42.3	1.0	0.14	0.53	0.98	95.5	0.97	0.95	0.93

Table 7.11: Accuracies using sector only look-up with all reference data (trimmed)

7.9.3 Sector and A-scan look-up

The previous 2-D representations take only a single peak value from all A-scan projections spanning each sector. As a result finer details within the A-scan group are hidden. One

method of preserving this information and making it more accessible is to convert the 3-D representation (figure 7.4.d) to a binary image (I) as follows:-

$$I_{(x,y)} = \begin{cases} 0 & \text{if } D^2_{x,y} \leq 0 \\ 1 & \text{otherwise} \end{cases}$$

An example of this procedure on the projection data for TB1PA1 (figure 7.4.d) produces the binary image illustrated in figure 7.6.a.

a) D^2 (after trimming) above $\chi^2_{(pc,0.95)}$ threshold
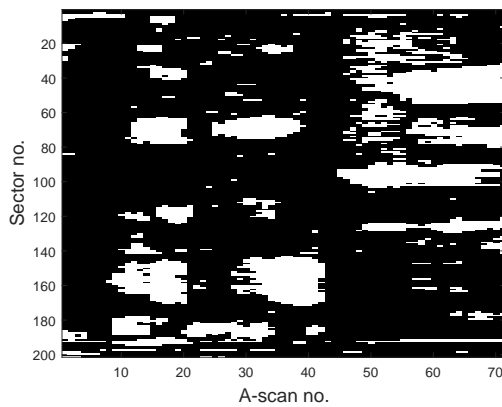
b) Ground truth mask
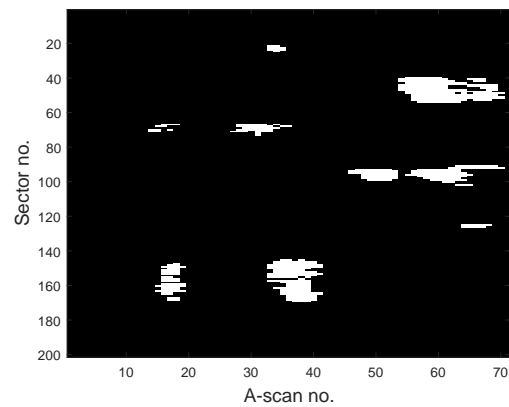
c) D^2 (after trimming) dB scale
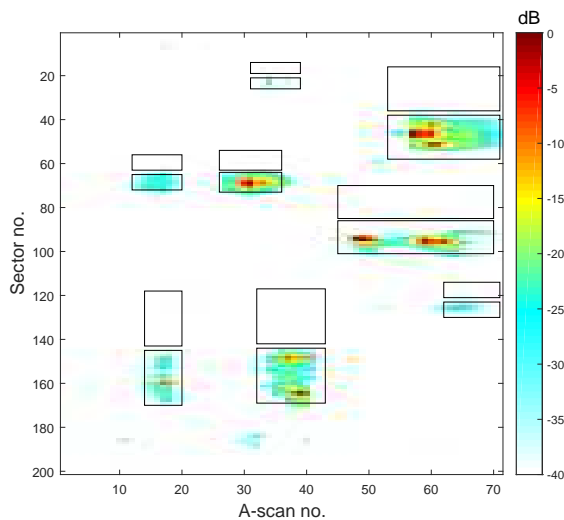
d) A-scan peaks dB scale
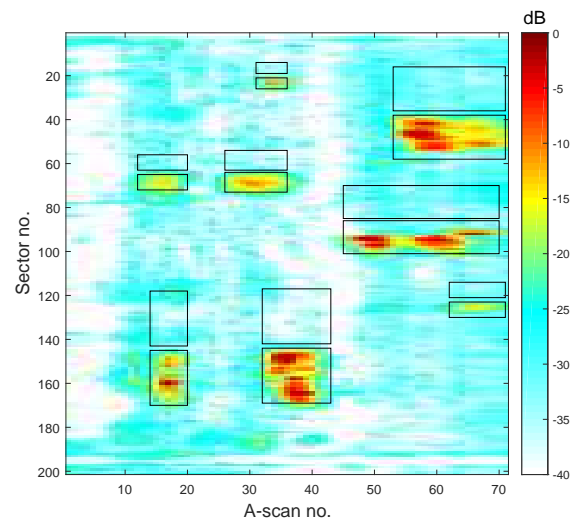
Figure 7.6: Example plan views of TB1PA1

Two observations now suggest that an improvement in classification is possible. Firstly

a comparison between figure 7.6.a and its ground truth 7.6.b indicates that the $\chi^2_{(pc,0.95)}$ thresholds are too low. A second comparison between the original projections on a dB scale, figure 7.6.c, with those for the equivalent plan view of A-scan peaks, figure 7.6.d, also on a dB scale, indicates a much improved signal to noise ratio due to the removal of the noisier principal components. Table 7.12 lists the D^2 SNRs after trimming the whole data set of 1816 observations. The image areas for the calculations, marked by the rectangles, correspond to those used previously for A-scan peaks and the manually selected training set. To ease comparisons these previous values are also listed.

The SNRs using the trimmed training set's Mahalanobis D^2 distance are, in all cases, significantly higher than those that use peak A-scan values and those using the manual training set. Improvements in SNR provide a stronger visual indication of anomalies but more importantly aid automatic classification.

It now becomes possible to compare the previous image thresholding techniques (Otsu, Kittler and Illingworth and Maximum Entropy) to automatically establish a threshold. Before this is done, and in keeping with the previous approach, a ROC analysis is used to determine the potential accuracies whilst maintaining a balance between sensitivity and specificity.

7.9.4 Full image ROC analysis

Table 7.13 compares the full image look-up accuracies for each test piece using the $\chi^2_{(pc,0.95)}$ confidence level alone and those following ROC analysis.

As expected the ROC threshold produces an improvement in specificity with a reduction in sensitivity. The improvements in the F1 statistic indicate a better discrimination. Although

Fault Instance	A-scan Peaks	SNRs (dB)	
		D^2	
		Manual (700)	Trimmed (1477)
Pore 1	23.67	24.9	43.1
Crack	31.25	38.1	51.0
Porosity I1	24.22	34.7	52.0
Inclusion 1	29.58	38.5	50.2
Pore 2	18.78	33.1	33.0
Inclusion 2 I1	35.82	54.4	61.2
Porosity I2	18.54	29.9	36.9
Inclusion2 I2	25.67	29.5	39.8

Table 7.12: SNR's comparisons

172

Test Piece	$\chi^2_{(pc,0.95)}$				ROC				
	Acc.(%)	sens.	spec.	F1	AUC	Acc.(%)	sens.	spec	F1
TB1PA1	83.9	1.00	0.83	0.33	0.99	95.5	0.96	0.96	0.63
TB1PA2	69.5	0.99	0.93	0.22	0.98	95.3	0.92	0.95	0.63
TB1PA3	84.7	0.99	0.84	0.35	0.99	95.1	0.96	0.95	0.62
TB1PA4	86.0	0.99	0.86	0.35	0.99	93.6	0.94	0.93	0.52
TB2PA1	81.3	0.92	0.81	0.28	0.95	95.9	0.90	0.96	0.63
TB2PA2	81.3	0.99	0.80	0.37	0.99	95.9	0.94	0.96	0.72
TB2PA3	89.0	0.94	0.90	0.36	0.97	94.4	0.90	0.95	0.52
TB2PA4	87.4	0.91	0.87	0.31	0.95	93.7	0.87	0.94	0.47
Averages	82.9	0.97	0.86	0.32	0.98	94.9	0.92	0.95	0.59

Table 7.13: ROC analysis (trimmed)

the F1 score gives some confidence in the ROC analysis the values are not particularly strong in this effect.

7.9.5 Automatic thresholding

ROC analysis does not provide a threshold a-priori. It is, however, now possible to investigate the thresholding techniques introduced in section 5.4. For each test piece table 7.14 gives a summary of threshold values, as a percentage of the total dynamic range, using each of the 3 thresholding methods (Kittler and Illingworth (KI), Otsu (O) and Maximum Entropy (ME) methods) alongside those from the previous ROC investigations.

Test Piece	R_{th}	KI_{th}	O_{th}	ME_{th}
TB1PA1	0.80	0.42	16.11	11.37
TB1PA2	0.72	0.45	12.95	10.19
TB1PA3	0.38	0.14	32.98	5.49
TB1PA4	0.34	0.18	16.50	7.45
TB2PA1	0.87	0.36	22.78	18.43
TB2PA2	0.19	0.08	14.91	4.31
TB2PA3	0.26	0.17	14.49	7.45
TB2PA4	0.69	0.44	14.55	9.80

Table 7.14: Projection thresholds as a % of total dynamic range

An immediate observation is that the thresholds using the Otsu and ME methods are, in comparison to the ROC thresholds, excessively high. This is thought to be due, in part, to the influence of large peaks in the Mahalanobis distances. Figure 7.7 illustrates the effect of each threshold for TB1PA1.

Test Piece	KI				Otsu				ME			
	Acc.(%)	sens.	spec.	F1	Acc.(%)	sens.	spec	F1	Acc.(%)	sens.	spec	F1
TB1PA1	91.8	0.98	0.92	0.50	97.0	0.26	1.00	0.39	97.4	0.36	0.99	0.52
TB1PA2	91.6	0.94	0.92	0.50	96.1	0.11	1.00	0.12	96.3	0.15	1.00	0.26
TB1PA3	90.1	0.99	0.90	0.45	95.9	0.02	1.00	0.04	97.2	0.33	1.00	0.50
TB1PA4	89.7	0.97	0.90	0.41	97.0	0.20	1.00	0.33	97.5	0.35	1.00	0.51
TB2PA1	90.5	0.90	0.90	0.42	96.8	0.17	1.00	0.30	96.9	0.21	1.00	0.36
TB2PA2	90.3	0.97	0.90	0.53	95.1	0.11	1.00	0.20	96.6	0.39	1.00	0.57
TB2PA3	92.2	0.91	0.92	0.44	97.2	0.16	1.00	0.28	97.7	0.29	1.00	0.46
TB2PA4	91.0	0.89	0.91	0.39	97.6	0.24	1.00	0.28	97.8	0.32	1.00	0.50
Averages	90.9	0.94	0.91	0.46	96.6	0.13	1.00	0.24	97.2	0.30	1.00	0.46

Table 7.15: ROC analysis (trimmed)
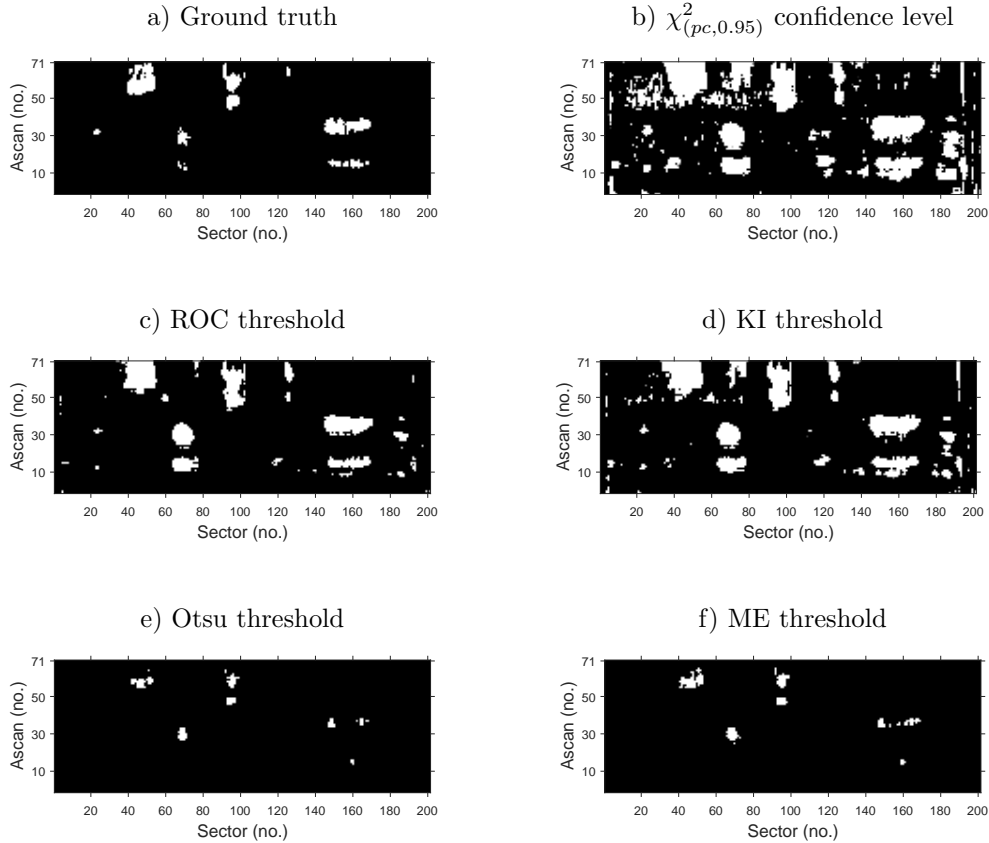
Figure 7.7: Comparison of thresholds on D^2 projections of TB1PA1

By comparison the KI thresholds appear more resilient to the peaks and more closely resemble those suggested by ROC analysis. A list of sensitivities, specificities and F1 scores for the KI thresholds are compared with those for the ROC in table 7.16. In general the sensitivity and specificity figures are similar to those obtained from the ROC analysis. The largest discrepancy with the ROC statistics is the sensitivity figure for TB2PA3. In this case the pore (fault F4, figure 4.5.b) is the only anomaly not detected.

For the inspector, an image such as 7.7.d, after KI thresholding, provides a visual indi-

Test Piece	ROC				KI			
	Acc.	sens.	spec.	F1	Acc.	sens.	spec.	F1
TB1PA1	95.5	0.96	0.96	0.63	91.8	0.98	0.92	0.50
TB1PA2	95.3	0.92	0.95	0.63	91.6	0.94	0.92	0.50
TB1PA3	95.1	0.96	0.95	0.62	90.1	0.99	0.90	0.45
TB1PA4	93.6	0.94	0.93	0.52	89.7	0.97	0.90	0.41
TB2PA1	95.9	0.90	0.96	0.63	90.5	0.90	0.90	0.42
TB2PA2	95.9	0.94	0.96	0.72	90.3	0.97	0.90	0.53
TB2PA3	94.4	0.90	0.95	0.52	92.2	0.91	0.92	0.44
TB2PA4	93.7	0.87	0.94	0.47	91.0	0.89	0.91	0.39
Averages	94.9	0.92	0.95	0.59	90.9	0.94	0.91	0.46

Table 7.16: Performance of KI threshold

cation of regions containing potential anomalies. Automation and further classification of anomalies requires a numerical list of these locations. Blob detection is one method of doing this. Figure 7.8 demonstrates the result of this step for TB1PA1. Each rectangle represent the boundary of a region above the KI threshold. This information may be presented to the inspector as a list. More importantly, as an extension to this work, A-scan data within each location may be subject to further analysis with the intent of classifying the particular anomaly.

Figure 7.8: Example blob detection (TB1PA1) above the KI threshold

175

7.9.6 Constant A-scan trimming using individual training sets

Reference data for projections using the constant A-scan orientation is, so far, the combination of all 8 test runs. As with the full sector orientation, the investigation now continues by limiting the original data set to that of the actual test under interrogation; that is, the reference set for TBn, PAn is the TBn, PAn data set itself. With the a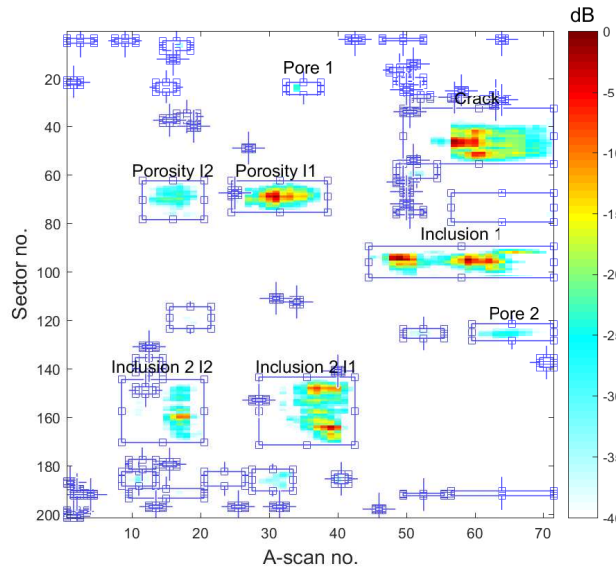vailable test sets this limits the size of the data set to a maximum of 227 observations. The intention is that this more closely represents the situation where no prior reference data is available. It is, once again, emphasised that in practice the weld length, and number of observations, will be much greater than the present test block limitation of 40 cm.

The reporting follows the previous pattern which considers sector only look up followed by sector and A-scan look-up. For each look-up method results are given using the $\chi^2_{(pc,0.95)}$ thresholds, before using ROC analysis to approximate an additional threshold to improve the balance between sensitivity and specificity. In the case of the combined sector and A-scan look-up a further investigation compares the ROC results with those from additional image thresholding of the Mahalanobis projections.

7.9.6.1 Sector only look up

Table 7.17 lists the results of sector only look-up using the individual trimmed reference sets. The evaluations are in terms of accuracy, sensitivity, specificity and the F1 score. For convenience corresponding results of the same projections using the full reference set are duplicated from table 7.11

Although both cases represent poor performance those for the individual case are, as ex-

Test Piece	Full $\chi^2_{(pc,0.95)}$				Individual $\chi^2_{(pc,0.95)}$			
	Acc.(%)	sens.	spec.	F1	Acc.(%)	sens.	spec.	F1
TB1PA1	40.8	1.0	0.13	0.52	35.3	1.0	0.04	0.50
TB1PA2	43.8	1.0	0.05	0.59	44.2	1.0	0.06	0.59
TB1PA3	46.9	1.0	0.21	0.55	36.3	1.0	0.05	0.51
TB1PA4	42.4	1.0	0.16	0.52	32.6	1.0	0.02	0.48
TB2PA1	36.8	1.0	0.10	0.49	33.3	1.0	0.05	0.47
TB2PA2	37.8	1.0	0.07	0.52	39.2	1.0	0.08	0.53
TB2PA3	49.8	1.0	0.25	0.57	34.2	1.0	0.01	0.50
TB2PA4	40.8	1.0	0.19	0.47	28.2	1.0	0.02	0.43
Averages	42.4	1.0	0.14	0.53	35.4	1.0	0.04	0.50

Table 7.17: Sector only look-up using full and individual data sets (trimmed)

pected, a degradation over those for the full reference set. For consistency with the previous study the projections are now subject to ROC analysis.

Results from the ROC analysis are listed in table 7.18. For comparison the equivalent results using the full data set (table 7.11) are duplicated. Despite the improvements the accuracy is inferior to that obtained using the larger data set.

Once again a conclusion here is that, although the proportion of anomalies is approximately the same in all reference sets (individual and full), the full data set contains more examples of anomaly free observations than do each of the individual reference sets. For example, after trimming the full data set reduces from 1816 observations to approximately 1500, whilst the individual sets reduce from 227 to around 140. With this type of application the indications are that provided the number of trimmed observations is sufficiently large then the Mahalanobis distance measure can tolerate more than the 30% contamination limit suggested by Jackson and Chen [103].

Of greater significance remains the expectation that in practice the length of an individual weld is likely to greatly exceed the 40 cm. limitation of the available test blocks. With longer welds the evidence suggests that the individual test set has the potential to support improved discrimination. However this can only be tested in the presence of longer test sets.

Test piece	Full ROC					Individual ROC				
	AUC	Acc.(%)	sens.	spec.	F1	AUC	Acc.(%)	sens.	spec.	F1
TB1PA1	0.99	95.5	0.94	0.96	0.92	0.87	80.6	0.85	0.78	0.74
TB1PA2	0.99	96.2	0.94	0.98	0.95	0.88	84.1	0.87	0.82	0.82
TB1PA3	0.97	95.2	0.99	0.94	0.93	0.97	92.3	0.88	0.94	0.88
TB1PA4	0.99	96.9	0.97	0.97	0.95	0.97	90.8	0.89	0.95	0.87
TB2PA1	0.99	96.1	0.98	0.95	0.94	0.88	82.4	0.87	0.80	0.75
TB2PA2	0.96	95.1	0.93	0.96	0.93	0.79	85.8	0.93	1.00	0.73
TB2PA3	0.99	94.6	1.00	0.92	0.93	0.97	94.1	0.93	0.95	0.91
TB2PA4	0.98	94.7	0.93	0.94	0.90	0.95	89.8	0.96	0.87	0.84
Averages	0.98	95.5	0.97	0.95	0.93	0.91	87.5	0.90	0.89	0.82

Table 7.18: ROC comparison sector only look-up full and individual data sets

7.9.6.2 Sector and A-scan (full image) look-up

With sector only look up the Mahalanobis distance provides a visual indication of sectors that potentially contain an anomaly. As there is no reliable way of determining a threshold automatic detection of smaller anomalies remains difficult. In the case of A-scan and

sector look up more information is available and although establishing a threshold remains a problem this can, as previously shown, be alleviated through image thresholding. This section now repeats the results for A-scan and sector look up, this time using the individual reference sets only.

Accuracy results for sector and A-scan look-up, using the full and individual trimmed reference data, are listed in table 7.19. The smaller individual training sets show a large improvement in accuracies. However some care is needed with this. Firstly the improvements do not represent particularly good values. In addition the potential improvements shown by the ROC analysis (table 7.20) are inferior to those using the full reference set (table 7.13).

The primary reason for the improvement, as evident from the specificity and sensitivity figures, suggests that this is due to a higher relative value for the $\chi^2_{(pc,0.95)}$ confidence limit. In the case of the full reference set there are more anomaly free examples which in turn produce a more accurate co-variance matrix and average value. Within this limitation the results continue to support the suggestion made using sector only look-up. That is, that provided the weld is of sufficient length, thereby providing more anomaly free observations, then the 30% contamination limit suggested by Jackson and Chen [103] is conservative. Consequently for many practical applications the individual data set will, after trimming, provide a suitable reference for reliable anomaly detection.

7.9.6.3   Automatic thresholding

Although the results for individual reference sets are not as robust as those using the full reference set it remains necessary, for completion, to present results for image thresholding.

Test piece	Full $\chi^2_{(pc,0.95)}$				Individual			
	Acc.(%)	sens.	spec.	F1	Acc.(%)	sens.	spec.	F1
TB1PA1	83.9	1.00	0.83	0.33	83.5	0.99	0.83	0.32
TB1PA2	69.5	0.99	0.93	0.22	84.7	0.98	0.84	0.36
TB1PA3	84.7	0.99	0.84	0.35	79.2	0.97	0.76	0.28
TB1PA4	86.0	0.99	0.86	0.35	79.4	0.95	0.79	0.26
TB2PA1	81.3	0.92	0.81	0.28	81.6	0.95	0.81	0.29
TB2PA2	81.3	0.99	0.80	0.37	87.5	0.92	0.87	0.45
TB2PA3	89.0	0.94	0.90	0.36	76.2	0.91	0.76	0.20
TB2PA4	87.4	0.91	0.87	0.31	73.7	0.98	0.73	0.19
Averages	82.9	0.96	0.86	0.32	80.7	0.96	0.80	0.29

Table 7.19: Sector and A-scan look-up using full and individual data sets (trimmed)

Test piece	Full ROC					Individual ROC				
	AUC	Acc.(%)	sens.	spec.	F1	AUC	Acc.(%)	sens.	spec.	F1
TB1PA1	0.99	95.5	0.96	0.96	0.63	0.97	94.7	0.95	0.92	0.57
TB1PA2	0.98	95.3	0.92	0.95	0.63	0.95	97.0	0.98	0.84	0.63
TB1PA3	0.99	95.1	0.96	0.95	0.62	0.96	94.2	0.94	0.93	0.56
TB1PA4	0.99	93.6	0.94	0.93	0.52	0.93	92.7	0.87	0.91	0.45
TB2PA1	0.95	95.9	0.90	0.96	0.63	0.93	97.6	0.95	0.81	0.67
TB2PA2	0.99	95.9	0.94	0.96	0.72	0.91	95.1	0.92	0.87	0.76
TB2PA3	0.97	94.4	0.90	0.95	0.52	0.92	97.8	0.76	0.98	0.66
TB2PA4	0.95	93.7	0.87	0.94	0.47	0.96	98.1	0.83	0.98	0.68
Averages	0.98	94.9	0.92	0.95	0.59	0.94	95.9	0.90	0.91	0.62

Table 7.20: Sector and A-scan look up individual trimmed data sets

Firstly table 7.21 lists the threshold values for the previous accuracy and ROC statistics (table 7.20). These are then compared with the thresholds obtained by subjecting each set of projections to thresholding using Kittler and Illingworth, Otsu and Maximum Entropy.

Once again the image thresholds illustrate that the Kittler and Illingworth thresholds more closely correlate with the accuracy (A_{th}) and ROC (R_{th}) thresholds. Table 7.22 provides an overview of the performance of the Kittler and Illingworth threshold. Results for both the very high Otsu and Maximum Entropy thresholds are omitted because they produce very poor sensitivity values.

On the whole the additional Kittler and Illingworth threshold produces a better accuracy than that indicated by the ROC analysis. However this is at the expense of sensitivity.

Test piece	Thresholds as % of dynamic range				
	R_{th}	KI_{th}	O_{th}	ME_{th}	Limit
TB1PA1	0.29	0.41	12.7	5.1	1.95
TB1PA2	0.98	0.32	59.02	5.1	2.04
TB1PA3	0.30	0.23	12.55	9.8	2.35
TB1PA4	0.32	0.15	23.5	6.67	2.25
TB2PA1	0.82	0.33	12.3	5.8	1.77
TB2PA2	0.64	0.005	30.3	2.7	0.84
TB2PA3	1.83	0.17	14.1	10.2	2.39
TB2PA4	0.52	0.002	29.6	5.1	1.5

Table 7.21: Comparison of threshold values for individual reference sets

Test piece	ROC					KI			
	AUC	Acc. (%)	sens.	spec.	F1	Acc (%)	sens.	spec.	F1
TB1PA1	0.97	94.7	0.95	0.92	0.57	95.6	0.83	0.96	0.60
TB1PA2	0.95	97.0	0.98	0.84	0.63	95.6	0.81	0.96	0.62
TB1PA3	0.96	94.2	0.94	0.93	0.56	93.5	0.93	0.94	0.54
TB1PA4	0.93	92.7	0.87	0.91	0.45	90.7	0.87	0.91	0.41
TB2PA1	0.93	97.6	0.95	0.81	0.67	96.2	0.85	0.97	0.64
TB2PA2	0.91	95.1	0.92	0.87	0.76	89.9	0.90	0.90	0.50
TB2PA3	0.92	97.8	0.76	0.98	0.66	94.2	0.89	0.94	0.50
TB2PA4	0.96	98.1	0.83	0.98	0.68	85.7	0.98	0.85	0.30
Averages	0.9	95.9	0.90	0.91	0.62	92.7	0.88	0.93	0.51

Table 7.22: Performance of KI threshold using individual reference sets

7.9.7 Summary of sector scan results

Anomaly detection of sectorial data using PCA has been investigated using two orientations. In the first approach, full sector projection, each sector is arranged as a single observation. This is analogous to many image recognition applications using PCA where an entire image is converted to a single vector. In this case the resulting vector has many dimensions (A-scans/sector $\times$ samples/A-scan) which very easily outnumber the maximum possible observations. Nevertheless it is demonstrated that with these HDLSS data sets anomaly detection remains possible.

The second orientation uses a constant A-scan for each observation. The number of dimensions is now limited to the number of samples/A-scan and is much reduced. Combining the full set of observations from all 8 test sets, the total number of observations, before trimming, now exceeds the number of dimensions.

For both orientations the approach has been to create reference sets using all available observations. Using PCA each observation is now projected against the eigenvectors for the entire data set. Any observation with a Mahalanobis distance greater than that specified by the $\chi^2_{(pc,0.95)}$ confidence limit is considered an outlier and is ejected from the original reference set. This process of trimming produces a new set of observations from which dominant outliers are removed. Repeating the PCA procedure gives a new set of eigenvectors that are now more robust to outliers and there is a greater probability of anomaly detection.

Results for full sector projections indicated accuracies above 90% with sensitivity and specificity value of 0.89 and above. One advantage of this approach is that, despite the high dimensionality, its operation is straightforward and efficient. It is limited to only

indicating sectors containing potential anomalies, giving no indication of their location within the sector. A ROC analysis of this data also indicates that the $\chi^2_{(pc,0.95)}$ threshold produces accuracies close to the optimums suggested by a ROC analysis.

The constant A-scan orientation is complicated by the fact that a set of eigenvectors is required for each A-scan. As suggested above these are of lower dimension than those for full sector projection and they permit two methods of look-up. Sector-only look-up provides an overview of sectors containing potential anomalies and is very similar to that for full sector projections. The alternative sector and A-scan look-up method provides both sector and A-scan information to locate a potential anomaly. Results from this analysis indicate that by itself the $\chi^2_{(pc,0.95)}$ confidence level tends to be too low. However a ROC analysis indicates the method has potentially greater accuracy than that for full sector projections. In addition the three dimensional results can be subject to image thresholding using conventional techniques. It is found that the Kittler and Illingworth method provides a threshold close to the optimum suggested using ROC analysis. With the available data this method produces more accurate results than those from full sector projections, at the occasional expense of a slight reduction in sensitivity, as shown by a comparison of tables 7.5 and 7.16.

Perhaps the most significant contribution of this section is that it establishes that automatic selection of a training set may be established through trimming and that the results are superior to obtaining a training set manually.

In addition to creating a reference data set from all 8 test sets, individual test sets were also used for each approach, the objective being to determine whether or not the data set itself provides sufficient information for anomaly detection. Individual test sets contain data from a maximum weld length of 40 cm. with a high density of anomalies. The conclusion here is that although the quality of the resulting covariance matrix and average value are suspect there is sufficient evidence to suggest that for longer weld lengths there will be sufficient data in the set. In particular it is concluded that for these type of defects the 30% contamination limit suggested by Jackson and Chen [103] is conservative.

7.10   TFM data sets

Unlike the available sectorial data the TFM images contain a dominating front wall reflection. It has previously been demonstrated (section 6.26) that one method of mitigating this effect is to standardise the data. As with the sectorial data it is necessary to evaluate

methods of establishing a training set automatically from the experimental data. In the case of the sectorial data two approaches were taken to establishing a reference set. Firstly, all data from all 8 test runs was combined. This was possible because the individual tests used identical set ups on two test pieces with the same dimensions. The results were then compared with those using individual test sets, limiting the reference set data from the particular test. In the case of the TFM tests, TFM3 has a weld cap making it significantly different from the TFM1 and TFM2. In addition both TFM1 and TFM2 contain only three small FBHs so that each has a large percentage of anomaly free reference data. Under these circumstances it was decided, for all of the TFM test pieces, to use reference data only from the specific test. As previously suggested this is more appropriate for practical situations where data from the single test set may be the only source of reference data. For the TFM tests trimming only applies to the individual test set.

The approach to trimming follows that for sector data with individual reference sets. That is, for each test piece, the whole data set is first subject to trimming. Once again the Mahalanobis distance is used to detect outliers, any outlying observation with a distance measure greater than the $\chi^2_{(pc,0.95)}$ value being ejected from the training set. The PCA operation is now repeated using the trimmed data set.

Previously it was established that of the three possible orientations the constant offset has by far the greatest potential. Although the constant depth orientation has some superiority for inspection within the bulk of the test piece it provides no cover for areas within a weld cap (or root). Additionally it loses reliability as the depth approaches either of the inner surfaces, tilting, or rocking, of images, as reported in section 6.26.2 and figure 6.18, being a particular problem for this orientation.

The previous chapter indicated that full image projections are unsuitable giving no indications of anomalies. For the record this chapter starts with a repetition of the full image projections but using trimmed data training sets. However, once again these proved to be unreliable so that most effort is towards investigation of the constant offset orientation.

7.10.1   Full image projections

Figure 7.9 provides a set of results for full image projections. Each case uses a training set acquired automatically by trimming the original data. Standardised data is used and the projections give no confidence in the ability of the approach to provide a reliable indication of anomalies. In view of this and the results of the similar investigation with manually derived training sets (figure 6.23) this orientation is given no further consideration. The

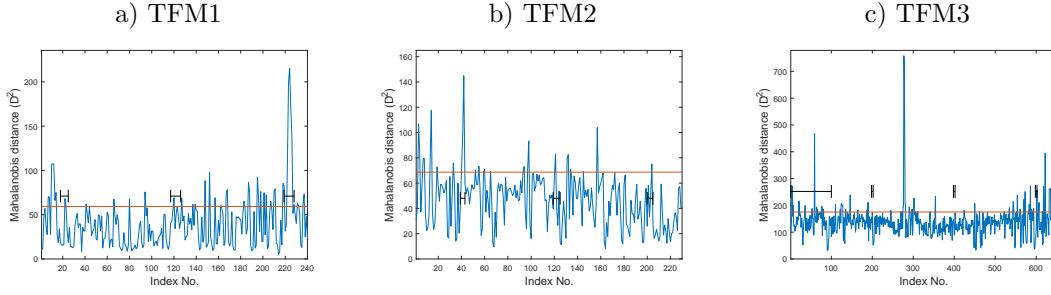remaining work considers only the constant offset orientation.

Figure 7.9: Full image projections with $\chi^2_{(0.95,PCs)}$ trimmed training sets

7.10.1.1 Constant offset projections

Acquisition of each training set is by ejecting any original observation with a projection greater than the $\chi^2_{(pc,0.95)}$ confidence level. This means that, unlike the manual training sets which were of fixed size, those obtained by automatic trimming are, for each offset, of a different size, even for the same test piece. An indication of the variation is listed below in table 7.23.

Test piece	Automatic trimming			Manual trimming
	Average	Min.	Max.	Fixed
TFM1	222	213	233	200
TFM2	213	201	221	200
TFM3	595	552	595	450

Table 7.23: Constant offset trimmed training set sizes

Figure 7.10 presents an overview of results for constant offset projections of each TFM data set. Comparison with the same results from the manually selected reference sets (figures 6.24 and 6.25) indicates some significant differences. In particular, with the manually selected training set, only TFM3 contained unusually high projections. After automatic training all data sets exhibit this characteristic to some extent. These differences tend to correspond to slices close to the beginning and end of each test run and are attributed to the method of training set selection. For the manual case the training set is a selection of slices considered to be free from anomaly. This naturally excludes slices containing the artificial anomalies. It did not exclude slices at the two ends of the data sets and, at the time, thought to be anomaly free, the exception to this being the slices at the end of TFM3 containing the SDH. Three of the plots in figure 7.10 (the exception being 7.10.c) show

some evidence of projections greater than background noise across all offsets corresponding to the end slices.
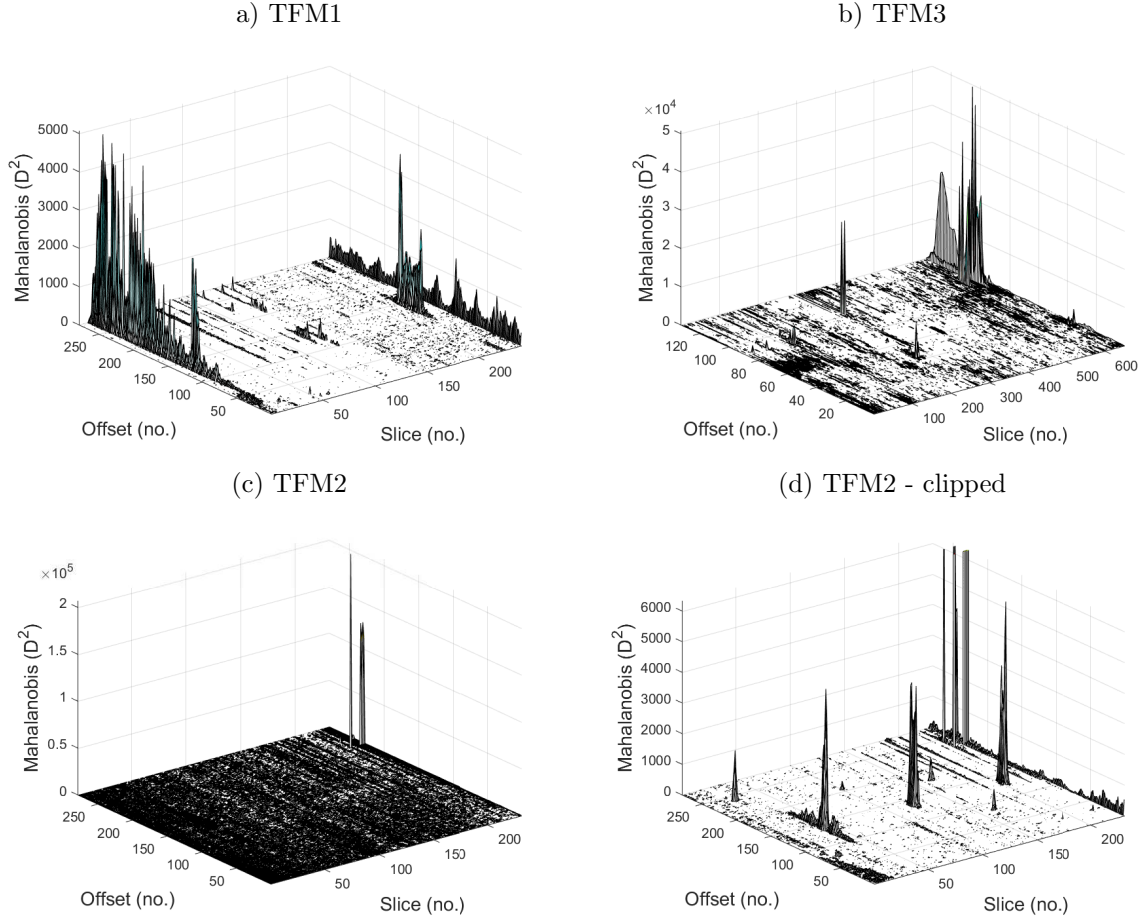
a) TFM1

b) TFM3

(c) TFM2

(d) TFM2 - clipped

Figure 7.10: Constant offset projections of TFM test pieces using trimmed training sets

These are most distinct in the case of TFM1. For this test piece figure 7.11 illustrates two constant offset images across all slices. At offset 50 the front wall is relatively flat across the entire length of the offset. For offset 250 the lower slice numbers, particularly around 5 to 20, indicate the position of the front wall to be less stable. This effect is due to a tilting in adjacent TFM images as reported in section 6.26.2 and figure 6.18. This therefore represents a real disturbance in the projections. In the case of the manual training set which included these observations the effect was masked out. The reason why this tilting effect exhibits itself most frequently at the two ends of each test run is not known.

On the whole the effect of the tilting is to create a 'wall' effect across all offsets of the
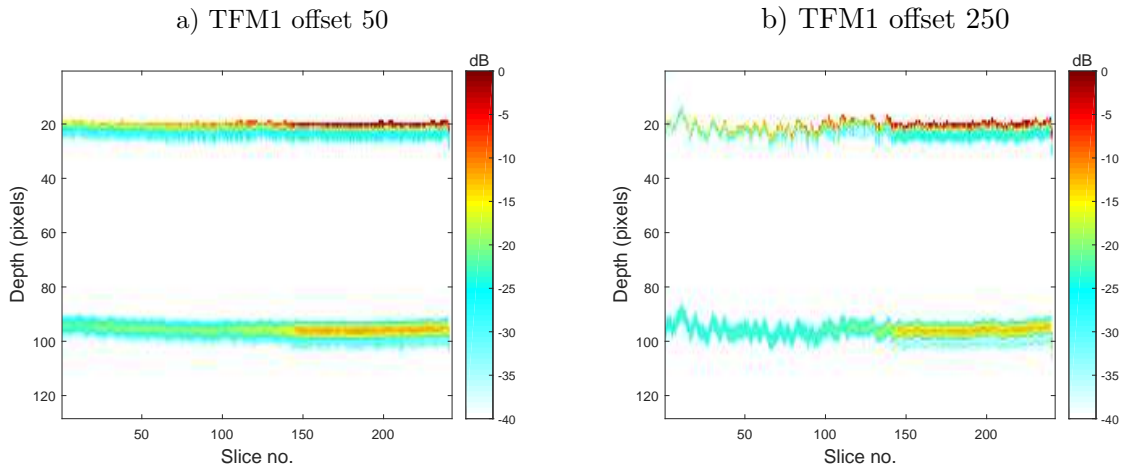
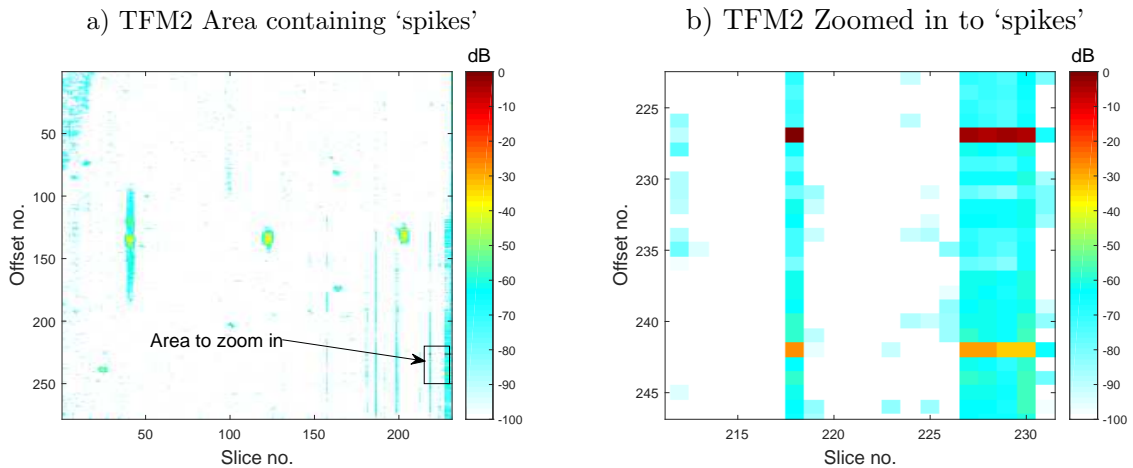Figure 7.11: TFM1 cross section at offset 50 and offset 250

Figure 7.12: TFM2 'spikes' - from figure 7.10.c

corresponding slices. However a further characteristic, within the same regions, is a small number of projections with unduly large magnitudes. The reason for these is not clear. They are particularly noticeable on the projections for TFM2 and TFM3 (figure 7.10.b and c). Despite their appearance on the illustrations and unlike a 'real' anomaly they tend to correspond to a small, in some cases a single, number of pixel points. For this reason they are unlikely to represent a real anomaly. To emphasise the small number of locations and erratic nature of these projections figure 7.12.a represents the plan view image of TFM2 on a dB scale. The two 'spikes', figure 7.10.c are the reason for the wide dynamic range of the image. To illustrate their size, in terms of pixel counts, figure 7.12.b zooms into the area they occupy. They are now seen to be present in only two single offset positions (227

and 242). At each offset one 'spike' occupies a single pixel (at slice 227) whilst the other occupies 4 pixels (across slices 227 to 230).

Whilst such abrupt changes reinforce the thought that they represent data processing artifacts rather than real anomalies, some consideration of the number of principal directions taken into account is required. Rejecting too many PCs may lead to important information being lost; equally using more PCs introduces more noise. At present the number of PCs is determined by the Kaiser stopping rule. Figure 7.14 repeats the projections of figure 7.10 but using only half the number of PCs specified by the Kaiser rule. For all projections this has reduced the magnitude (anomaly and background). The effect of this on each anomaly's SNR is now used as a measure of the effectiveness of this for detection.

7.10.1.2   Signal to noise ratio measurements

Using the Kaiser PCs, figure 7.13 illustrates images of the projections for each TFM test piece on the same dB scale. Areas indicated by the rectangles are those used for the SNR measurements of each anomaly, table 7.25. These are the same areas used for the SNR measurements of the manual training set (section 6.31.3) and for convenience the previous results are duplicated in table 7.24. After discussing these this section considers the effect of a reduction in the number of PCs used (table 7.26).

As well as enabling early investigations a purpose of the manual training set, selected to exclude anomalous locations, was to provide a set of results by which those from any automatically derived training set could be compared. A comparison of the two tables does in fact demonstrate this. With two exceptions, F1 in TFM1 and the unintended anomaly F5 in TFM3, all SNRs from the manually selected training set exceed those from the trimmed training set. That said all anomalies have some presence.

Despite the comparative inferiority of the trimmed training set the results provide some confidence in the approach. This is particularly significant as there is no manual intervention. The exercise has also demonstrated a danger with manual selection not exhibited by automatic selection. This refers to the condition whereby the manual selection actually contained observations not representative of the background and discussed in section 7.10.1.1. Although masking, due to anomalous observations in the manual training set, produces projections giving a clearer indication of the intended anomalies, those using the trimmed training set are a more accurate representation of the actual TFM images. The higher SNRs of the manual data set are due to the masking of background noise. In view of the more accurate representation of the actual TFM images and the values of the SNRs

measured, trimming is demonstrated to be a viable way of automatically creating a training set without any knowledge of the actual ultrasonic set up conditions.

Further effects of masking, by the manual training set, is the absence of projections representing the tilting of some TFM images. The manual training set excluded only slices containing known anomalies. For the trimmed training set any observation containing a potential outlier is excluded. This is likely to include observations containing evidence of tilt. Although not strictly an anomaly these are not masked by the trimmed training set.

Whilst table 6.15 lists the SNRs of each anomaly using the Kaiser stopping rule, table 7.26 lists the same SNRs using only half of these. In almost all cases the SNRs are reduced. Experiments using other ranges of PCs indicate a similar outcome. Using fewer PCs does suppress the unexplained spikes. As this exercise demonstrates that there is also a reduction in anomaly SNR, the Kaiser stopping rule will continue to be used. A further motive for use of Kaiser is that it provides a method of automatically determining the number of PCs to use.

After establishing a plan view of each test piece below the front wall the next stage is to evaluate methods of automatically detecting the location of anomalies. Following the previous examples this is attempted through slice only look-up and constant offset with slice look-up.

Test Piece	Anomaly SNRs (dB)				
	F1	F2	F3	F4	F5
TFM1	42.1	26.1	43.2	-	-
TFM2	45.2	45.1	48.1	-	-
TFM3	22.7	19.7	25.0	21.2	34.5

Table 7.25: Trimmed training set

Test Piece	Anomaly SNRs (dB)				
	F1	F2	F3	F4	F5
TFM1	24.9	24.6	23.7	-	-
TFM2	36.8	38.6	37.6	-	-
TFM3	21.1	18.2	25.7	22.9	34.8

Table 7.26: Trimmed training set SNRs (KaiserPCs/2)

Test Piece	Anomaly SNRs (dB)				
	F1	F2	F3	F4	F5
TFM1	22.9	35.1	56.2	-	-
TFM2	46.3	53.3	57.1	-	-
TFM3	51.6	27.0	30.8	24.1	14.5

Table 7.24: Manual training set

7.10.1.3 Slice only look-up

Slice only look-up takes the maximum distance across all offsets for each slice to provide a graphic summary of potential anomalies over the whole test piece. The corresponding
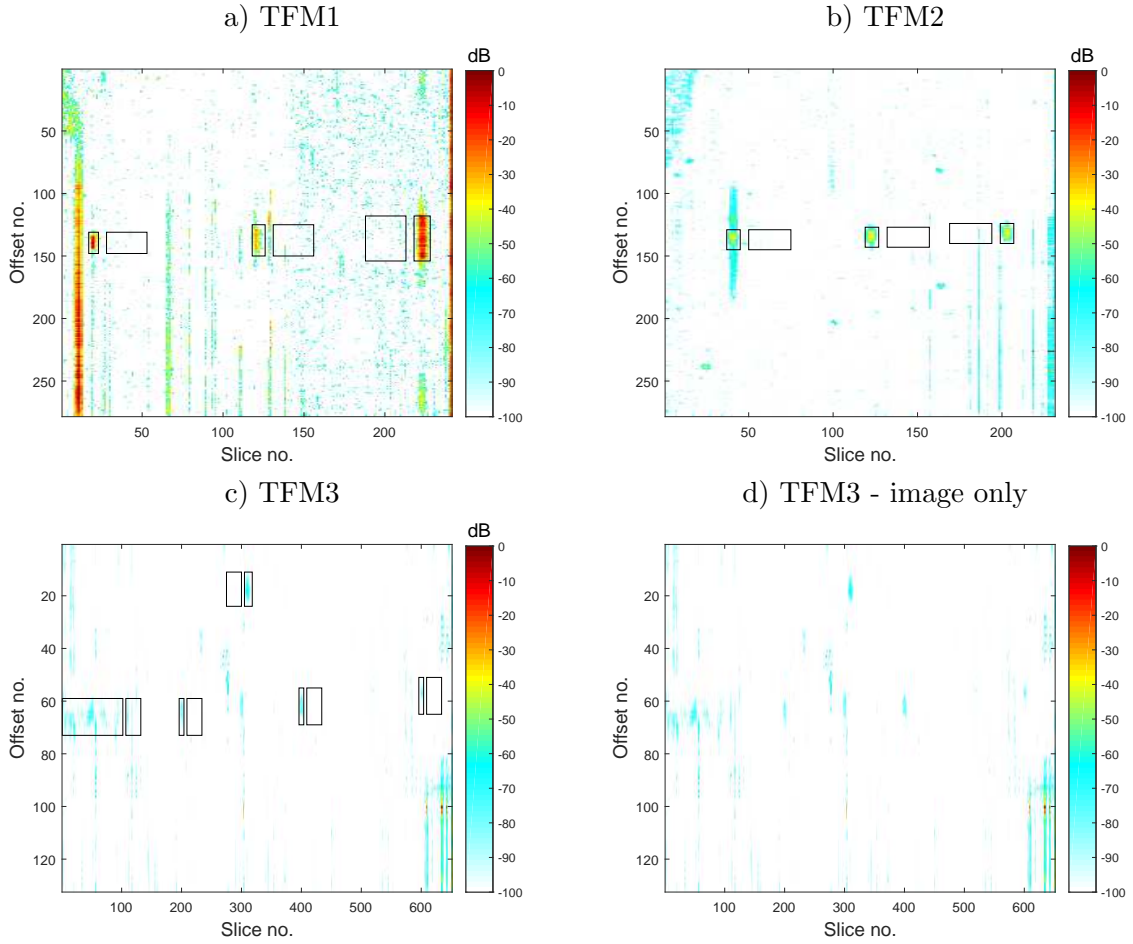
Figure 7.13: Constant offset TFM images with outlines for SNR calculations

plots are illustrated in figure 7.15, along with the respective ground truths. In comparison to the equivalent plots for the manually selected reference sets (figure 6.30) these indicate more false positives. Of most significance is the fact that the SDH in TFM3 is not well pronounced.

Test Piece	χ^2				ROC					
	Acc (%)	Sens.	Spec.	F1	R_m	AUC	Acc (%)	Sens.	Spec.	F1
TFM1	14.1	1.0	0.08	0.12	1.88	0.85	87.1	1.0	0.86	0.49
TFM2	12.6	1.0	0.02	0.12	4.00	0.87	87.0	1.0	0.87	0.51
TFM3	21.2	1.0	0.02	0.32	1.94	0.76	74.0	0.86	0.72	0.57

Table 7.27: Constant offset $\chi^2_{(pc,0.95)}$ and ROC accuracies - slice only look-up
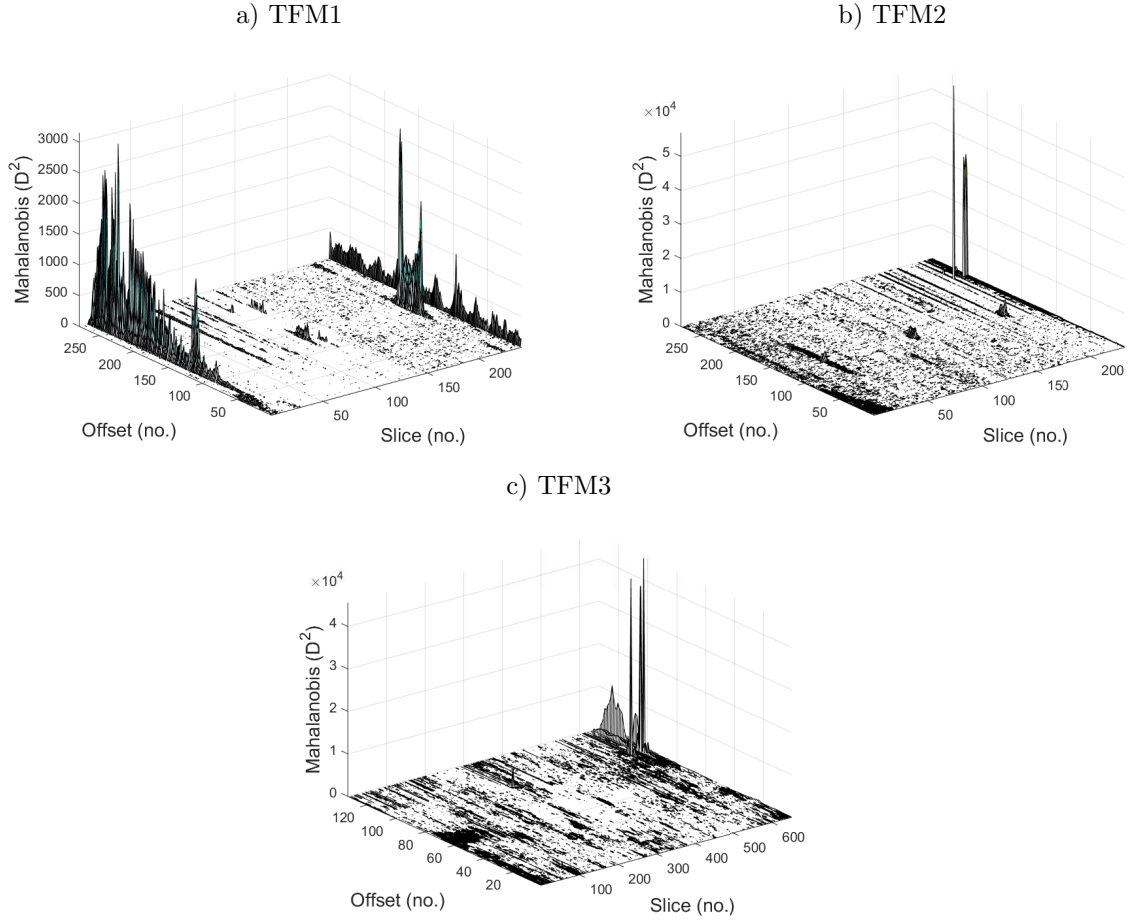
a) TFM1   b) TFM2

c) TFM3

Figure 7.14: Constant offset TFM images using reduced number of PCS

7.10.1.4   Slice and offset look-up

Slice and offset projections, as illustrated in figure 7.10, already have the respective $\chi^2_{(pc,0.95)}$ confidence limit subtracted from each projection. Any value greater than zero is a potential anomaly so that its conversion to a binary image provides a mask for comparison with the respective ground truth binary template. Figure 7.16 illustrates both the ground truth and the binary conversion of each projection. A logical comparison between each pair provides similarity statistics for each test piece.

The $\chi^2_{(pc,0.95)}$ columns of table 7.28 list the results of this comparison in terms of classification accuracy, sensitivity and specificity. Overall these results are a considerable improvement over those for slice only look-up. Although the F1 scores are poor, results of the ROC analysis show very little change in sensitivity with a larger improvement in
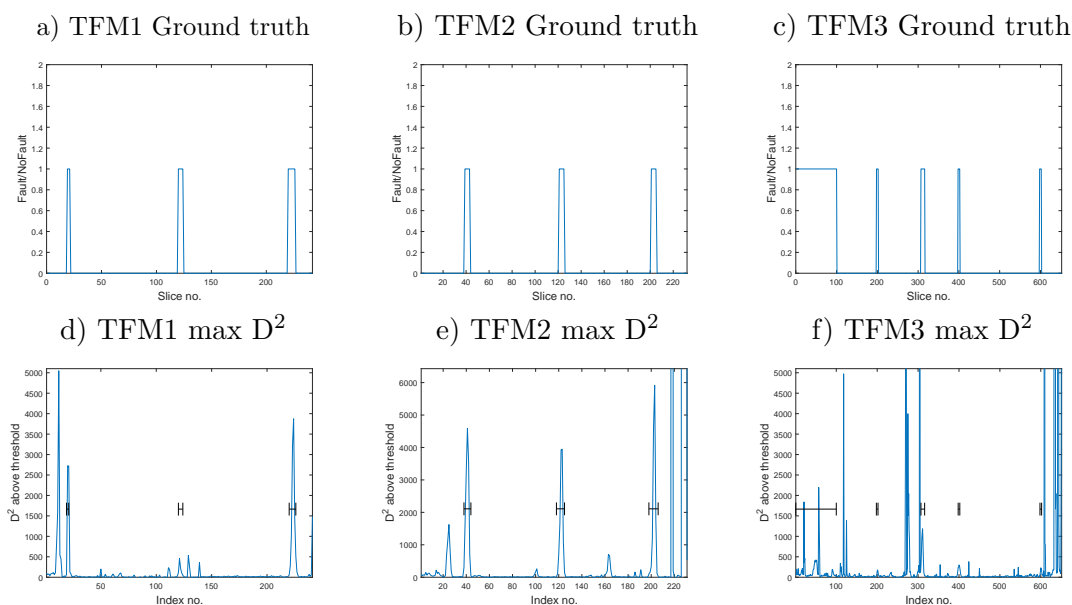
Figure 7.15: Slice only look-up

specificities, suggesting that the $\chi^2_{(pc,0.95)}$ confidence level provides a reliable threshold for anomaly detection. A further consideration of the binary images in figure 7.16 confirms this to be the case. In particular all of the $\chi^2_{(pc,0.95)}$ images provide some evidence of each anomaly, suggesting that the poor F1 statistics are in fact due to the number of false positives.

As ROC analysis provides a threshold only in the presence of a ground truth each original set of D^2 projections is now subject to each of the previous thresholding techniques. Table 7.30 lists the results of this exercise for all test pieces. For completeness resulting images for each test piece are illustrated at the end of this chapter (figures 7.18, 7.19 and 7.20).

Inspection of the sensitivity and specificity figures for the Otsu and ME thresholds indicates each to be far too high. For TFM1 and TFM2 the KI threshold produces results close to those recommended by the ROC analysis. Unfortunately this is not the case for TFM3 where the KI threshold is also high.

7.10.1.5 Thresholding

The large discrepancy between the thresholding techniques and in particular the poor performance of the Otsu and ME methods warrants investigation. Some discussion of the
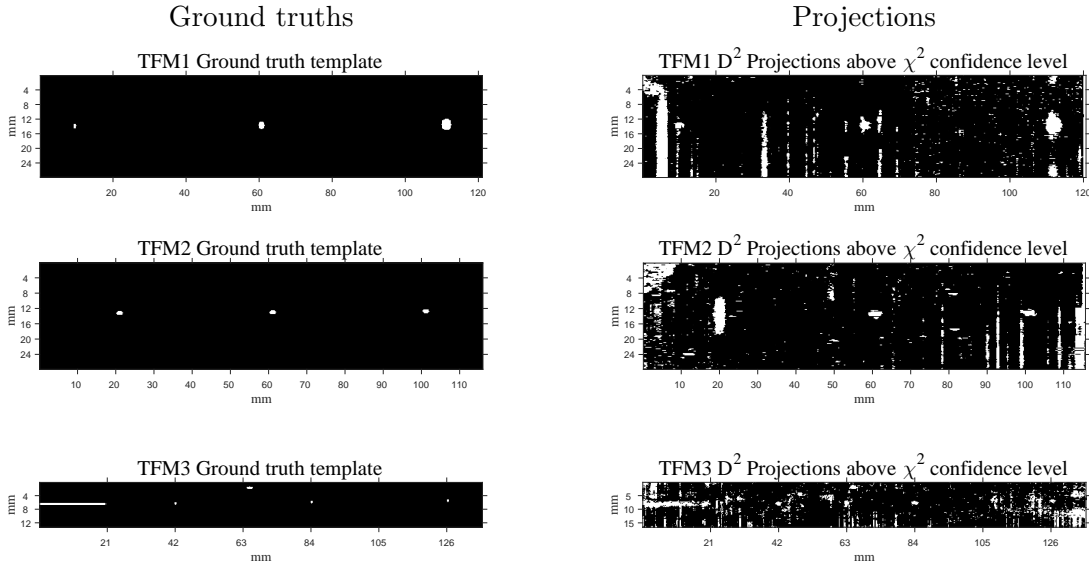
Figure 7.16: Ground truths and projections above the $\chi^2_{(pc,0.95)}$ limit for TFM1, 2 and 3

small number of spuriously large projection distances evident in all projections, and most pronounced in the case of TFM2 (figure 7.10), was given in section 7.10.1.1. To determine how detrimental these might be to thresholding the exercise was repeated after suppressing the largest projections. For each test piece the limiting value is set to 3 standard deviations above the mean.

Statistics using these limited D^2 projections are listed in table 7.31. Overall these were found to be little different from the originals, table 7.30. Although the Otsu and ME threshold values are reduced they remain far too high. This indicates that spuriously high projections alone are not responsible for the poor results and that further explanation for this deficiency is required, particularly as Sezgin and Sakur [49] demonstrate all three techniques to outperform many other thresholding methods.

A primary requirement of both the Otsu and ME methods is that foreground objects and background must be of a similar size [53], [52]. For all TFM test pieces this is not the case, the number of foreground objects being outweighed by the background. A second requirement is for the projections to be bimodal. In the case of these projections the majority of gray values are very low with only a small number, representing anomalies, being large. As an example the histogram for the TFM1's projections, figure 7.17, indicates a unimodal rather than bimodal characteristic. Histograms with similar sparse patterns were exhibited in data used by Drobchenko et al. [110] who reported similar limitations of

Test piece	$\chi^2_{(pc,0.95)}$				ROC					
	Acc (%)	Sens.	Spec.	F1	R_m	AUC	Acc (%)	Sens.	Spec.	F1
TFM1	89.6	0.99	0.89	0.06	1.36	0.98	96.3	0.98	0.96	0.14
TFM2	90.3	0.91	0.90	0.03	1.13	0.94	95.6	0.90	0.96	0.04
TFM3	86.7	0.86	0.87	0.12	1.01	0.93	85.1	0.88	0.85	0.11

Table 7.28: Trimmed training set confusion statistics - full image look-up

Test piece	$\chi^2_{(pc,0.95)}$				ROC					
	Acc (%)	Sens.	Spec.	F1	R_m	AUC	Acc (%)	Sens.	Spec.	F1
TFM1	92.0	1.0	0.92	0.07	1.86	0.99	98.5	1.00	0.98	0.28
TFM2	92.0	0.94	0.92	0.03	1.16	0.96	95.2	0.94	0.95	0.05
TFM3	85.5	0.94	0.85	0.12	1.77	0.96	94.5	0.92	0.95	0.26

Table 7.29: Manual training set confusion statistics - (duplicated from table 6.12)

Test Piece	Method	R_m	Accuracy (%)	Sensitivity	Specificity	F1
TFM1	D^2	1	89.6	0.99	0.89	0.06
	D^2_{ROC}	1.36	96.3	0.98	0.96	0.14
	KI	1.45	96.8	0.95	0.97	0.16
	Otsu	20.34	99.5	0.19	0.99	0.21
	ME	14.42	99.5	0.31	0.99	0.21
TFM2	D^2	1	94.2	0.90	0.94	0.03
	D^2_{ROC}	1.33	95.6	0.89	0.96	0.04
	KI	1.42	97.0	0.84	0.97	0.08
	Otsu	1277	99.9	0.00	1.0	0.00
	ME	85	99.8	0.03	1.0	0.06
TFM3	D^2	1	86.4	0.87	0.87	0.12
	D^2_{ROC}	1.01	85.1	0.86	0.88	0.11
	KI	1.6	95.8	0.56	0.96	0.22
	Otsu	17954	99.2	0.00	1.0	0.00
	ME	936	98.9	0.00	1.0	0.0

Table 7.30: Classification statistic results for different thresholds

the Otsu and ME methods. In further agreement with these results they also reported the KI method to perform better. For unimodal histograms Drobchenko et al. [110] evaluated an alternative technique due to Rosin [111]. This is based on finding the corner in a unimodal histogram. Here the conclusion was, once again, that the KI method produced the better results.

Test Piece	Method	R_m	Accuracy (%)	Sensitivity	Specificity	F1
TFM1	D^2	1	69.9	0.99	0.70	0.02
	D^2_{ROC}	1.11	96.3	0.98	0.96	0.14
	KI	1.49	97.4	0.94	0.98	0.16
	Otsu	3.02	98.8	0.25	0.99	0.11
	ME	2.05	98.3	0.35	0.99	0.12
TFM2	D^2	1	72.2	0.96	0.72	0.01
	D^2_{ROC}	1.13	94.2	0.86	0.94	0.04
	KI	1.42	97.0	0.80	0.97	0.07
	Otsu	21.22	99.9	0.40	1.0	0.46
	ME	7.86	99.8	0.40	1.0	0.46
TFM3	D^2	1	88.3	0.839	0.83	0.10
	D^2_{ROC}	1.01	85.1	0.88	0.85	0.11
	KI	1.6	95.9	0.56	0.96	0.22
	Otsu	248.95	98.9	0.00	1.0	0.00
	ME	70.67	98.8	0.0	0.99	0.0

Table 7.31: Confusion statistic results for 3σ limited projections

7.10.2 Comparison with the manually selected training set

The expectation is that in addition to indicating a greater potential for discrimination, the manual training set will also provide an indicative upper limit of performance. A comparison of results from the manual training set and those using the automatically acquired trimmed training set is now made on the basis of the respective SNRs (tables 7.24 and 7.25) and confusion statistics (tables 6.14 and 7.30). For convenience these results are duplicated here in tables 7.32, 7.33 and 7.34. To simplify the threshold comparisons, results for the Otsu and ME methods are omitted.

7.10.2.1 Signal-to-noise ratios

With three exceptions the SNR comparisons of the two training sets follow a consistent pattern, with those for the manual training set being higher than the respective value of the trimmed case. The exceptions (TFM1:F1, and TFM3:F1,F5) are notably different.

Leaving aside TFM3:F1, for the time being, the remaining two exceptions show a considerable improvement over the SNRs of the manual training set. Although, overall, the respective SNRs of the manual training set tend to be the higher, those of the trimmed training set are comparable. This demonstrates that in these instances trimming is an effective method of improving the performance of PCA.
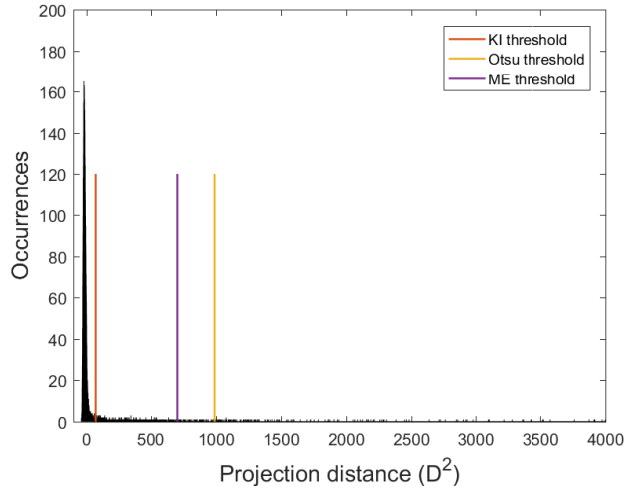
Figure 7.17: TFM1 - Histogram of D^2 projections and thresholds

In the case of TFM3 the respective SNRs for F1 represent an exception. Using the manual training set this anomaly has considerable prominence but is subdued in the trimmed case. This is possibly due to the method of training set acquisition. Observations for the manual training set are selected as slices deemed to be free from anomaly. In the case of the trimmed training set it is outlying offsets, not slices, that are rejected. One consequence of this is that in the manual training set observations containing the anomaly are completely absent. This is not the case for trimming where some observations may contain residual evidence of an anomaly. Nonetheless, and despite the presence of a weld cap, the SNR of F1, using the trimmed training set, is not insignificant.

Test Piece	Anomaly SNRs (dB)				
	F1	F2	F3	F4	F5
TFM1	22.9	35.1	56.2	-	-
TFM2	46.3	53.3	57.1	-	-
TFM3	51.6	27.0	30.8	24.1	14.5

Table 7.32: Manual training set

Test Piece	Anomaly SNRs (dB)				
	F1	F2	F3	F4	F5
TFM1	42.1	26.1	43.2	-	-
TFM2	45.2	45.1	48.1	-	-
TFM3	22.7	19.7	25.0	21.2	34.5

Table 7.33: Trimmed training set

7.10.2.2 Sensitivity and specificity

As suggested an increase in an anomaly's SNR is likely to improve its probability of detection. However in the context of this work, a more significant test is based on the actual ability of the technique to discriminate an anomaly from its background. In table 7.34 the confusion statistics are compiled into three measures, accuracy, sensitivity and specificity.

Whilst a low value of accuracy represents a poor classification a high value does not necessarily represent a good classification (section 5.11.1). Accuracy values need to be used with caution and are presented primarily for comparisons. More meaningful measures of discrimination are given in the form of sensitivity (indicating a high proportion of true positives with few false negatives) and specificity (indicating a high number of true negatives with few false positives). In short a sensitivity of 1 and specificity of 1 represent a complete match between test data and its ground truth.

With few exceptions the respective sensitivities and specificities for the two training sets are similar but in most cases those for the manual training set are slightly higher. The more significant differences are for TFM3. This is thought to be due to the relatively low values of SNR (in particular that for F1, the 20 mm. deep SDH). However, this does not necessarily mean that anomalies are not detected. It rather implies that some pixel values, corresponding to a known anomaly, may register as false positives. Similarly a low specificity implies a number of false positives. Examples of this are illustrated in the binary images for each test piece (figures 7.18, 7.19 and 7.20).

Test Piece	Method	Manual training sets				Trimmed training sets			
		R_m	Acc.(%)	Sens.	Spec.	R_m	Acc.(%)	Sens.	Spec.
TFM1	D^2	1	92.0	1.0	0.92	1	89.6	0.99	0.89
	ROC	1.86	98.4	0.99	0.98	1.36	96.3	0.98	0.96
	KI	1.44	97.7	0.99	0.98	1.45	96.8	0.95	0.97
TFM2	D^2	1	92.0	0.94	0.92	1	94.2	0.90	0.94
	ROC	1.16	95.3	0.94	0.95	1.33	95.6	0.89	0.96
	KI	1.39	97.0	0.88	0.97	1.42	97.0	0.84	0.97
TFM3	D^2	1	85.5	0.94	0.85	1	86.4	0.87	0.87
	ROC	1.77	94.5	0.92	0.95	1.01	85.1	0.86	0.88
	KI	1.6	93.7	0.92	0.94	1.6	95.8	0.56	0.96

Table 7.34: Classification statistics for different thresholds

For all test pieces the greatest sensitivity is recorded when the threshold value is that of the confidence limit alone ($R_m = 1$). ROC analysis tends to indicate an increase in specificity with only a slight reduction in sensitivity. With the exception of TFM3 application of KI thresholding to the D^2 projections produces results comparable to those of the ROC threshold. For TFM3 the KI threshold is likely to be too high. At face value classification results for TFM3 give some concern, both the sensitivity and specificity figures being low. This test piece is, however, particularly demanding. Not only are the results from an immersion test but the anomalies are also beneath a weld cap. This, and the fact that the three SBHs represent relatively weak reflectors, results in a binary image, figure 7.20.a, containing numerous false positives. It will be shown that, although the application of blob

detection to such an image reveals many locations erroneously identified as anomalous, the application of a few simple rules can greatly reduced the number of false positives whilst still giving indications of all real anomalies.

7.10.3   Blob detection

As indicated (table 7.34) thresholding the D^2 projections using the χ^2 confidence level produces the highest sensitivities whilst the KI threshold produces results most closely resembling those from ROC analysis. After thresholding a list of areas containing a potential anomaly can be produced using blob detection. Figures 7.21.a, c and e illustrate the result of applying the χ^2 thresholds. Here each region of interconnected blobs is bounded by a rectangle. The blob detection procedure provides the location of the top left corner, width, height and area of each rectangle. The minimum size of a rectangle is an isolated foreground pixel. Where the region contains more than one pixel the width and height of the rectangle is determined by the aspect ratio of the respective blob. One consequence of this is that the actual number of foreground pixels is generally smaller than the area of its enclosing rectangle.

Although blob detection provides a list of potential anomalies there are clearly far too many false indications. Figures 7.22.a, c and e illustrate a similar set of results using the KI threshold. This produces far fewer false positives, as suggested by the respective specificity figures. Despite the low sensitivity for TFM3 all five anomalies continue to be highlighted. The largest anomaly, F1, is identified in a series of four rectangles (rather than one) along the length of the SDH.

7.10.3.1   Image post processing

Many of the blobs detected after χ^2 confidence level or KI thresholding are in fact individual pixels. Although they may be indicative of a true anomaly they are more likely to result from noise or one of the spuriously high spikes reported earlier (section 7.10.1.1). In general these single pixel rectangles only serve to obfuscate the images. Their removal from the image gives greater emphasis to the larger rectangles which are more likely to be representative of true anomalous regions. A second source of obfuscation is the effect of 'tilting' in the original TFM images (section 7.10.1.1). The effect of this is to produce columns of foreground pixels particularly noticeable in TFM1 and TFM2 (figures 7.18.a and 7.19.a). These are known not to be caused by anomalies, although once again, there is some possibility that they may hide such an instance.

Figures 7.21.b, d, f and 7.22.b, d, f illustrate the results of blob detection after making two simplifications to their respective originals. The post processing steps include identification of overly large vertical bars, areas outside terminating at the top or bottom boundary and areas that are below a certain size (9 pixels in the examples). This exercise is done only to show the potential for post processing. Confusion statistics continue to be used for measuring and comparing classification performance.

7.10.4   Summary of TFM results

Detection of anomalies within TFM images has been investigated using only the constant offset orientation. At an early stage full image projections were found not to be encouraging and were not pursued. The primary purpose of this section is, however, to investigate the use of trimming as a means of obtaining a reference set for PCA. In this context the anticipation is, that if the reference set is sufficiently representative of the anomaly free background, then it will provide an accurate low rank sub space.

Results from the previous chapter, where the reference set was selected manually with a-priori information on the location of anomalies, are expected to provide an upper limit on which the results from automatic trimming may be compared. With the exception of TFM3 the results, in terms of anomaly SNRs, sensitivities and specificities were comparable indicating that automatic trimming is effective. In the case of the exception, TFM3, it was found that despite the relatively poorer figures, in particular the sensitivity, it remained possible to detect some presence of each anomaly. Although KI thresholding provides some improvement in overall sensitivity and specificity use of the χ^2 confidence level itself, with some post processing of the resulting image, is satisfactory. This produces a greater number of false positives than using KI thresholding. However in these cases the false indications do not present a major detraction.

In further mitigation of the apparently poor results for TFM3 it is postulated that differences in the acquisition of the two training sets are, in part, responsible for apparently superior performance of the manual training set. In practice the training set must be acquired by a data driven approach. Acquisition of the manually acquired training set uses slices deemed to be free from anomalies. For automatic training set selection the full image projections use slice only trimming. This approach was found not to be effective. For the more successful constant offset approach trimming is by ejecting offsets with unusual projections. This is different from the approach used for manual selection. It is also difficult to manually select offset data deemed to be free from anomaly. Therefore it might be argued

that the comparison between manual and trimmed data sets is not on a true basis.

Overall, PCA using an automatically derived training set through trimming is, however, demonstrated to be an effective method of anomaly detection in TFM images. Of more significance the major contribution of this, and the previous, chapter is to demonstrate the successful creation of a plane view image of subsurface anomalies through a weld cap from an ultrasonic immersion test.

7.11   Chapter summary

This chapter has examined the application of PCA to both sectorial data and TFM images. In both cases the recommended method of analysis is similar. In the case of sectorial data the recommendation is to use constant angle A-scans as the reference set (one for each angle). After trimming each training set is used to create an image using sector and A-scan look-up. Compared with using the peak values of each A-scan, anomalies in the resulting image have a significantly improved SNR, increasing their probability of detection.

For TFM images the recommendation is to use constant offsets for the reference set (one for each offset). After trimming each training set is used to create an image using slice and offset look-up. This is found to produce a plan view not otherwise possible without 'physical' removal of the front wall. Alternative attempts to remove the front wall entail making estimates of the wall's position. This is error prone with a requirement to make many assumptions. Using the PCA approach, with standardised data removal of the wall, is implicit.

In both cases creating a training set by automatic trimming effectively improves the robustness of PCA. This was demonstrated by earlier comparisons using untrimmed (all observations) and trimmed training sets. With untrimmed data one of the major limitations of PCA, its sensitivity to outliers, was demonstrated. Alternative methods of improving the robustness of PCA is to use a robust covariance matrix. Recent approaches to this include Principal Component Pursuit (PCP). Like the trimming approach used here PCP is data driven. An overview and evaluation of the approach is given in the next chapter.
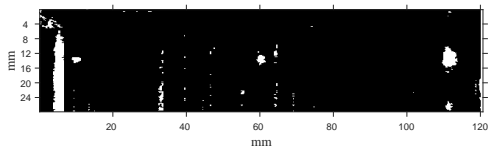
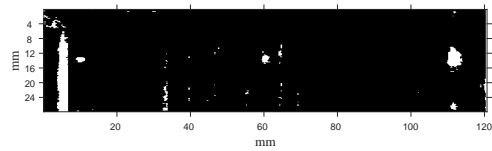a) D^2 above $\chi^2_{(pc,0.95)}$ Confidence level
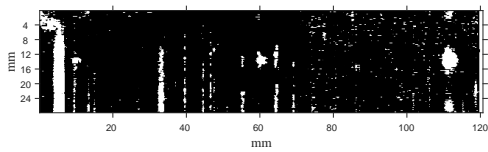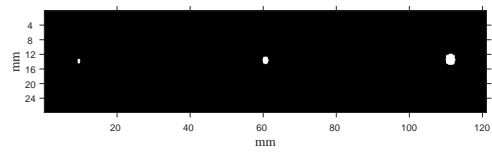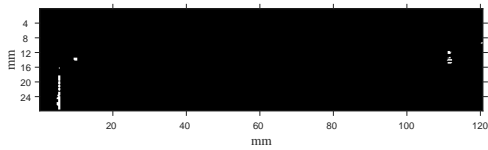
b) Ground truth

c) ROC threshold

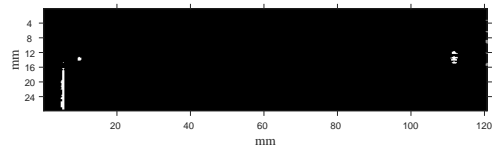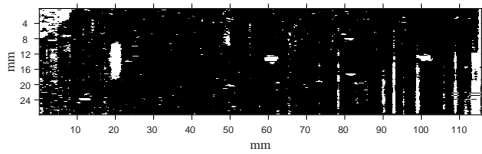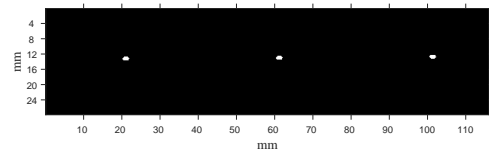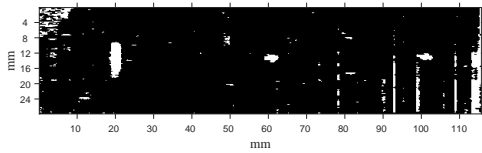d) KI threshold

e) Otsu threshold

f) ME threshold

Figure 7.18: TFM1 D^2 above $\chi^2_{(pc,0.95)}$ confidence level with additional thresholding

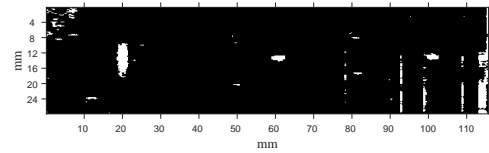a) D^2 above $\chi^2_{(pc,0.95)}$ confidence level

b) Ground Truth

c) ROC threshold

d) KI threshold

e) Otsu threshold

f) ME threshold

Figure 7.19: TFM2 D^2 above $\chi^2_{(pc,0.95)}$ confidence level with additional thresholding

a) D^2 above $\chi^2_{(pc,0.95)}$ confidence level
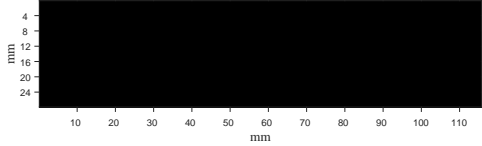
b) Ground truth
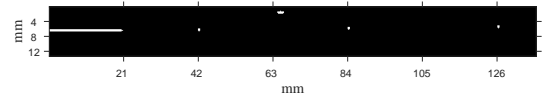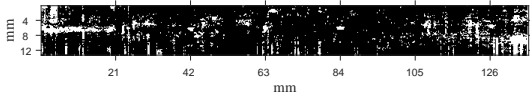
c) ROC threshold
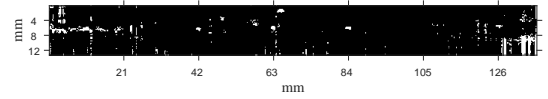
d) KI threshold

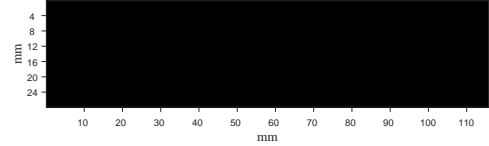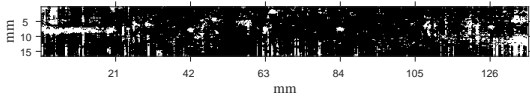e) Otsu threshold
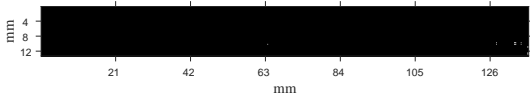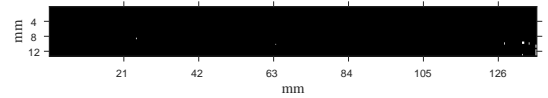
f) ME threshold

Figure 7.20: TFM3 D^2 above $\chi^2_{(pc,0.95)}$ confidence level with additional thresholding

a) TFM1 original
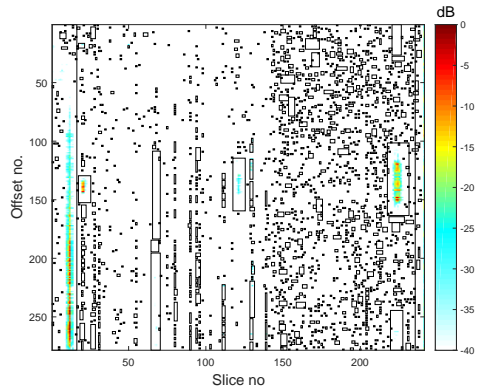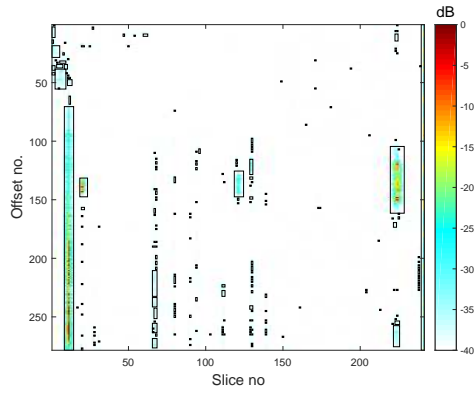
b) TFM1 after post processing
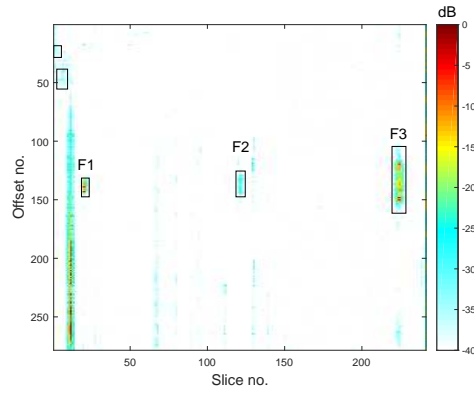
c) TFM2 original
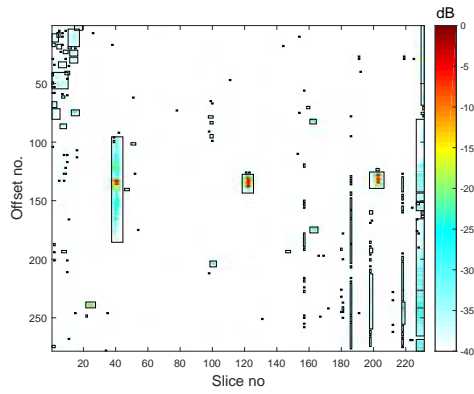
d) TFM2 after post processing
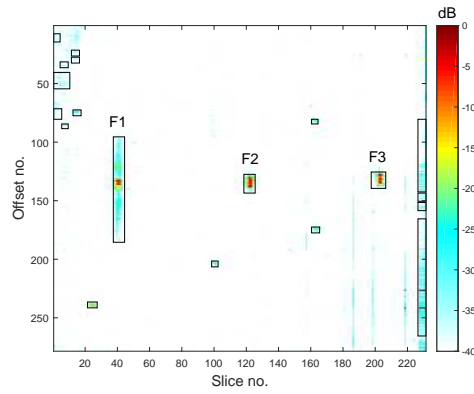
e) TFM3 original
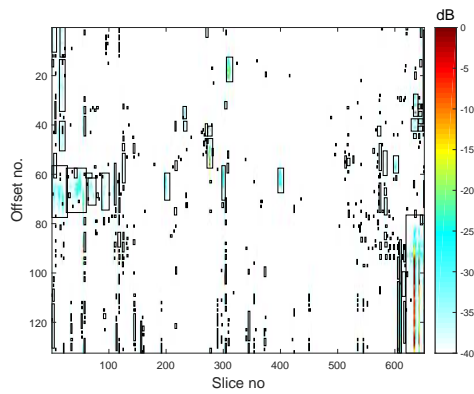
f) TFM3 after post processing

Figure 7.21: Potential fault locations by blob detection after χ^2 thresholding

202
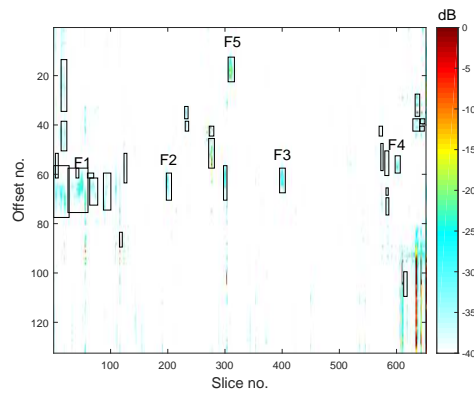
Figure 7.22: Potential fault locations by blob detection after KI thresholding

Chapter 8

Robust PCA

Robust Principal Component Analysis (RPCA) is a generic term that refers to various methods of overcoming the problems of outliers. In this context trimming, used 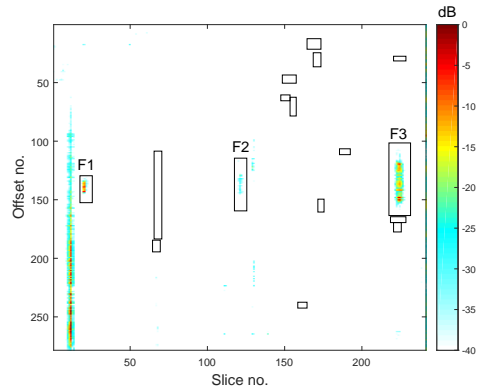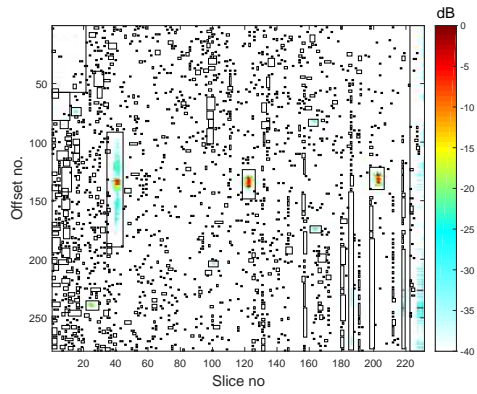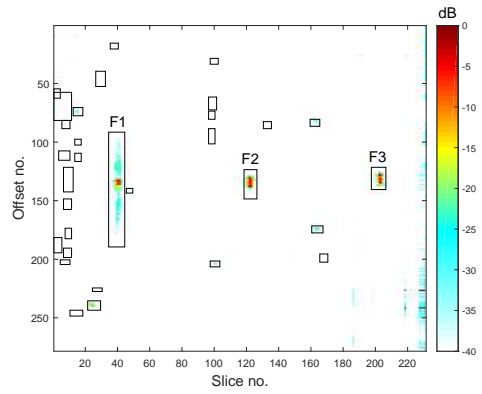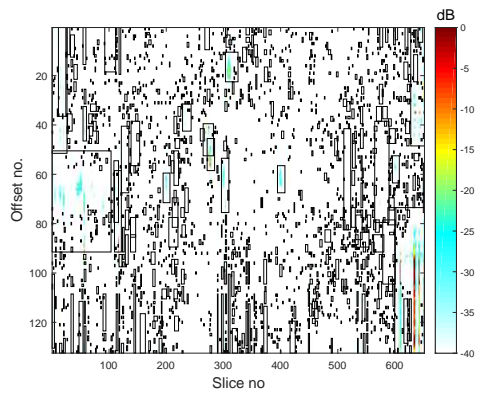in the previous chapter, may be included in this category. This chapter discusses an alternative, more formal approach, based on Principal Component Pursuit (PCP) which is currently receiving considerable attention [107], [112], [113]. The original application of this approach relates to video surveillance where the objective is to separate a moving foreground (intrusion or anomaly) from a static background. In the case of a set of ultrasonic images the video information is provided by 'flipping' through a full sequence of images. Regions that are largely static represent the fault free background. Moving objects, or foreground features, represent potential anomalies.

Common to many of these methods the full sequence of images is first converted into a single matrix where each column represents a single image. PCP is now used to decompose the matrix into one matrix containing static or background information and a second matrix containing any moving or foreground information. As each column of the background matrix contains relatively few differences the matrix is of low rank. The foreground matrix which typically contains many zeros (no changes) is sparse. In the context of ultrasonic inspection each matrix may be used for anomaly detection. For example, the low rank subspace may be used as a basis for projection of an unknown image, large distances being an indication of an anomaly. Alternatively, where there is no intrusion, the sparse matrix will be empty. If an intrusion does occur the respective columns of the sparse matrix will contain unusually high variation.

After presenting background information on PCP the chapter considers methods of apply-

ing the technique to anomaly detection. Results are presented and comments made on the application of the technique.

8.1   Limitations of PCA

PCA aims to minimise the squared reconstruction error (Pearson [114]) or maximise the variance of the projected data (Hotelling [115]). Traditionally this is achieved by constructing a low rank subspace approximation to the zero mean training data that is optimal in a least squares sense [116], [117]. This is equivalent to minimising the Euclidean or L2 norm of the vectors. It has many properties making it extremely popular. In particular L2 expressions are well understood, derivatives are easily obtained and a unique closed form solution is always obtained. However it is also commonly known [118], [119] that when using a least squares estimate larger elements have a greater influence on the result so that the technique is not robust.

The previous chapter gave an illustrative description of trimming. Here a similar illustrative example is given of robust PCA. To further the comparison the same two dimensional data set (table 8.1) is used as an example of the PCP algorithm [107]. As a reminder, the particular data set is chosen so that each dimension has zero mean. A plot of the original data set is denoted by the blue crosses ($\times$) in figure 8.1.a and the outlier is clearly visible at co-ordinate (10,0).

Without the outlier the points represent a straight line, and a Singular Value Decomposition (SVD) of the data set would result in a single eigenvector. However, due to the outlier, the SVD operation results in two eigenvectors in the orientations indicated by the two solid orthogonal lines. A projection of each original data point, using the first principal component, results in a set of data points represented by the red circles. The presence of the outlier means that there is now a reconstruction error (dashed lines) between the original data points and the projection. In the absence of the outlier the projection would be onto the original data points and there would be no reconstruction error. In the data trimming example the outlier is detected and ejected from the original data set so that a second application of SVD results in a more accurate low rank approximation.

PCP is robust against such outliers. That is to say, it attempts to produce a low rank subspace without ejecting outliers from the data set. In figure 8.1.b this is done using the technique to be presented. It should be noted that after allowing for rounding errors the original matrix, X, is equal to the sum of the low rank, L, and sparse, S, representations. On

205

the graph the blue crosses represent the original data set, each point of the corresponding low rank representation (L) being denoted by an asterisk (*). On the graph these have been normalised (zero-centred) before plotting. Using SVD with the low rank data results in a single eigenvector (solid line), and a projection of the data points results in zero reconstruction error.

X1	=	-6	-5	-4	-3	-2	**10**	0	1	2	3	4
X2	=	-5	-4	-3	-2	-1	**0**	1	2	3	4	5

Table 8.1: 2 dimensional data with added outlier (X)

X1	=	-3.37	-3.37	-3.02	-2.01	-1.00	0.02	0.00	1.00	2.00	3.00	3.37
X2	=	-3.35	-3.35	-3.00	-2.00	-1.00	0.02	0.00	0.99	1.99	2.98	3.35

Table 8.2: Low rank representation of 2 dimensional data with outlier (L)

X1	=	-2.62	-1.62	-0.98	-0.99	-0.99	9.98	0.00	0.00	0.00	0.00	0.62
X2	=	-1.64	-0.64	0.00	0.00	0.00	-0.02	1.00	1.00	1.01	1.02	1.64

Table 8.3: Sparse representation of 2 dimensional data with outlier (S)

a) Original data set   b) Low rank decomposition
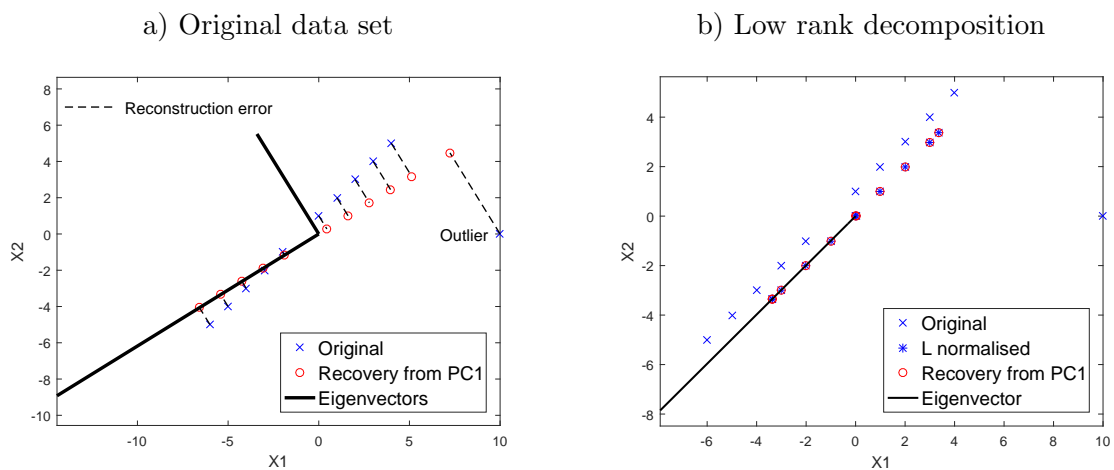
Figure 8.1: Example of low rank PCA

8.2 RPCA background

In a general sense RPCA refers to any method of obtaining a low rank subspace in the presence of outliers. As previously suggested active research and applications of RPCA

are most evident in the field of intelligent video surveillance, Cermeño et al. [120]. In this context the image background represents a low-rank subspace with any movement representing an intrusion or sparse outlier. This approach has now been used in other areas. For example, in a networked computer system where parameters such as network traffic, memory usage, processing and disk activity are monitored, any unusual outliers could indicate attempts of financial fraud or cyber attack. The literature also indicates that such techniques have not been applied to ultrasonic anomaly detection. Here there is a need to detect variations within a sequence of indexed images along the direction of a weld. A low rank subspace represents an anomaly free background, with any localised anomaly constituting a foreground object which moves into and out of the scene representing a sparse outlier. As the low-rank matrix models the continuous, possibly slowly changing, background it becomes the training set for conventional PCA.

To improve robustness in the presence of outliers the previous chapter used trimming to provide a more accurate representation of the anomaly free data conditions. Such methods were first proposed in the early 1970's [97] and since then other approaches have been investigated. In addition to the Mahalanobis and MVE methods discussed earlier, these include the application of weight functions to the mean and outer products of the covariance matrix (Huber [119] and Cambell [121]), whilst Fischler and Bowles [122] developed a technique based on random sampling. Generally these methods are computationally impractical for high dimensional data such as sector scans and TFM images.

8.3   RPCA overview

More recently various versions of RPCA have been developed and a comprehensive classification of the techniques is given by Bouwmans et al. [123]. Many of these methods rely on replacing the L2 by the L1 norm, a particular advantage being that the L1 norm is less sensitive to outliers. Unfortunately, many of the advantages of the L2 norm are lost and the solution is not necessarily unique. One of the earliest reports of using the L1-norm are due to Baccini et al. [124] who used a heuristic estimate, whilst Ke and Kanade [125] [126] proposed convex programming methods. Ding et al. [127] report a method of combining the merits of L1 and L2. Whilst successful at suppressing the effects of outliers this method is highly dependent on the dimension of the low rank subspace and being a power method [128] it is, for a large dimensional space, time consuming to converge. Due to its robustness to outliers research interest in the use of L1 PCA continues [129]. Unfortunately in terms of minimising the reconstruction error or maximising the variance of the projected

data, none of these methods guarantee strong performance under a broad set of conditions.

8.4   Low rank and sparse representations

A further property of the L1 norm, not so far discussed, is that it produces a sparse solution [130]. 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. Sparse systems are loosely coupled and their decomposition provides an efficient description in terms of simpler components [131]. Consequently, an alternative interpretation of Robust PCA, whereby the matrix X is decomposed into low rank (L) and simpler sparse (S) components, is:-

$$X = L + S \tag{8.1}$$

Previously this work has used data trimming to improve the robustness of PCA. If, instead, SVD is performed on the above low rank matrix L, with sparse corruptions removed:-

$$L = U\Sigma V^T \tag{8.2}$$

then the columns of U form an orthonormal basis (eigenvectors) for the true subspace. A comprehensive review of many methods for solving the RPCA problem, where the data matrix is split into a low rank (background) matrix L and a sparse (foreground) matrix S, is given by Bouwmans et al. [112]. Amongst theses are Principal Component Pursuit (RPCA-PCP), Candès et al. [107], Proximal Gradient and Continuation [132] and Rank Sparsity Incoherence [131].

In spite of the numerous proposals for formulating RPCA, Bouwmans et al. [112] suggest that, to date, no algorithm has emerged that is able to simultaneously address all of the key challenges of real video data processing. They do, however, conclude that PCP presently provides the most representative formulation for the problem. For this work the immediate objective is to investigate, for the first time, RPCA as a method of detecting defects in a sequence of indexed ultrasonic images. Unlike a continuous video sequence, the ultrasonic data is available as a batch of images representing a single weld, providing some simplification. At this stage there is, therefore, no reason to compare the apparent idiosyncracies of the alternative approaches to RPCA. Consequently the PCP formulation outlined by Candès et al. [107] is to be used. A further motivation for this is the availability (through Stanford University) of the open source Template for First Order Conic Solvers

(TFOCS) software ([133]). TFOCS provides a set of Matlab building blocks that can be used to solve a variety of convex models, particularly those requiring sparse recovery. Practice with this package provided background familiarity with the procedure. However for the more 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 (Laptev [134]). In this case the optimisation is provided using the Alternating Direction Method of Multipliers (ADMM) procedure [135]. Simulations [136] indicate that the ADMM is slightly more efficient, in terms of computation time and number of iterations, than TFOCS.

8.5 Overview of PCP

Before presenting results a brief outline of the PCP algorithm, used in all examples, is given. In practice the operations are performed using the ADMM algorithm (section 8.4). The objective is to split the matrix into a low rank and sparse components. From seminal works [107], [131] recovery of a low rank matrix is obtained by the following method:-

$$
\begin{aligned}
\text{minimise} \quad & \|L\|_* + \lambda\|S\|_1 \\
\text{subject} \quad \text{to} \quad & X = L + S
\end{aligned}
\tag{8.3}
$$

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 can be chosen by experimenting over a few iterations. Alternatively it can be obtained from:-

$$
\lambda = \frac{1}{\sqrt{max(r,c)}}
\tag{8.4}
$$

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

In fact equation 8.4 is proven to hold (section 1.5 and theorem 1.1 of [107]) with a probability of almost 1 providing:-

- the rank of L is small:-

$$
\leq n/((log(n)))^2 \quad where \quad n = max(r,c)
$$

- and S is reasonably sparse:-

$$\leq (n^2)$$

A further caveat is that S has a uniformly distributed sparsity pattern. The significant advantage of equation 8.4 is that no tuning of the parameter is required.

8.6   Visualisation of the L and S matrices

A simple example is now used to provide insight into the use of the low rank and sparse matrices. The original data set, figure 8.2.a, is a matrix, M, containing all data from a fixed offset of TFM2. Here offset 133 is selected as it intercepts all three FBHs. The slice number corresponds to the TFM image from which the specified (offset) column of data is taken. Each row of the column corresponds to a pixel depth. The complete data set contains two continuous peaks across approximate pixel depths 20 and 95. These correspond to the front and back walls respectively. Between these are three isolated peaks representing each of the three FBHs drilled to different depths.

Figures 8.2.b and c illustrate the same data after decomposition of M into its low rank, L, and sparse, S, representations respectively. In this case there is a clear separation of the background (low rank) data and the foreground (sparse data) containing anomalies. Visually the low rank (L) matrix contains mainly the front and back wall, the three FBHs being strongly attenuated. Ideally the sparse (S) matrix will contain evidence only of the three peaks which will be completely absent from L. However some variations in the front and back walls are also evident. These are due to slight differences in the location and orientation of the front and back wall from adjacent TFM images. The 'floor' of the sparse data image also indicates some slight variations. By definition most cells of the sparse matrix will contain zero with small variations resulting from computational precision; in this context, on all figures, 'white' areas correspond to cell values less than zero.

For this application the decompositions can be used in a number of ways. Through the low rank matrix it is possible to approximate a basis for the anomaly free subspace (equation 8.2), thus replacing the data sets formerly obtained through manual selection or trimming. Alternatively the sparse matrix may be used to detect outliers directly. This can be by inspection of individual images or by sequencing through a series of sparse images. In this case anomalies are evident by making a dynamic transition into and out of the sparse image. This separation is equivalent to one of the earliest reported foreground detection methods [137]. Although, in this case, the sparse representation is dominated by variations
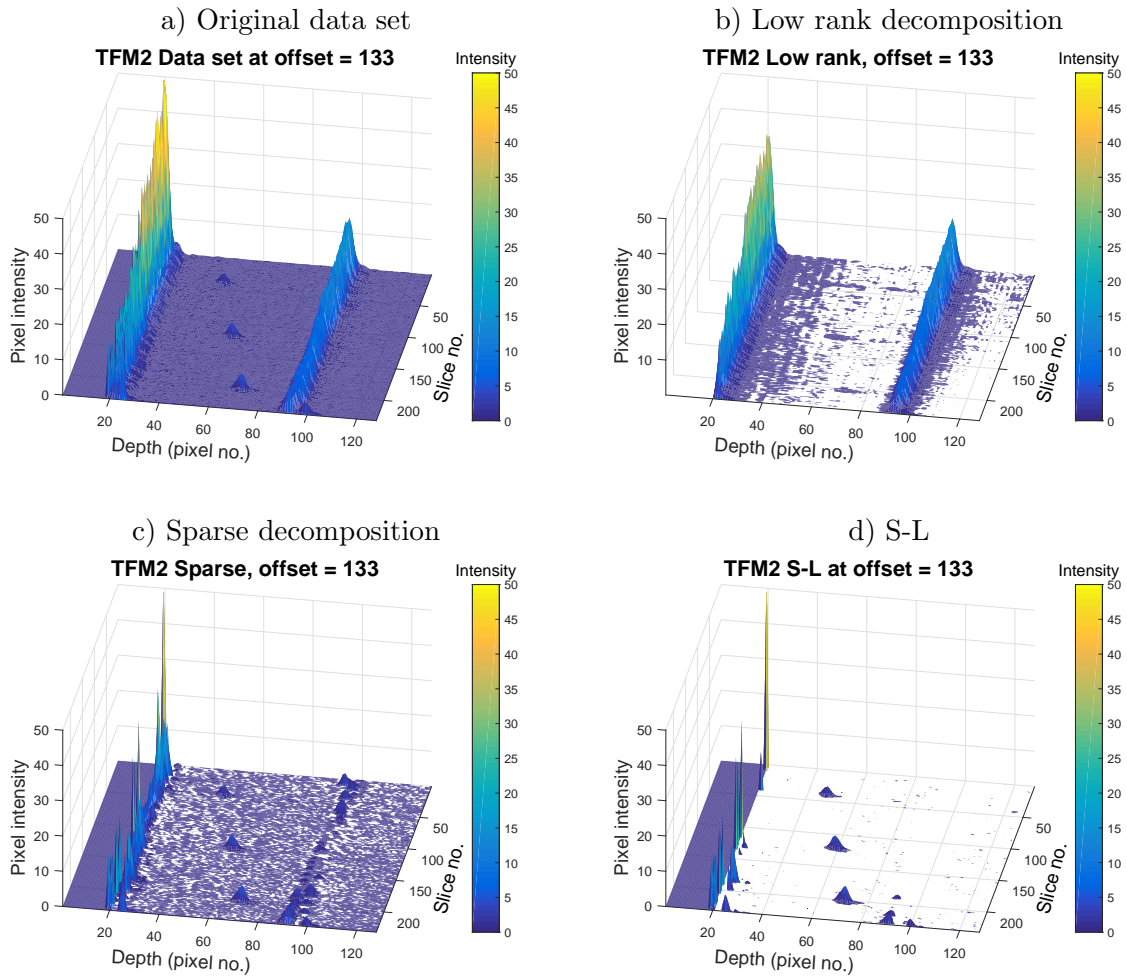
Figure 8.2: Low rank and sparse decomposition using TFM2, offset 133

in the front wall, this can be mitigated by subtracting the low rank representation (figure 8.2.d).

8.7 Application of PCP to ultrasonic data

The remaining sections of this chapter present the findings of investigations based on PCP. As a representation of the 'outlier free' background the low rank matrix is now used, for anomaly detection, by conventional PCA.

One immediate issue regarding the technique is the computational time required for the iterations of the ADMM algorithm. This is due mainly to the SVD computations needed

to compute each iteration of the low rank matrix, the SVD operation with computational complexity $O(n^3)$ consuming most of the time. Ideally the iterations terminate on meeting the constraint L + S = X (equation 8.3). In many cases it is not practical to pursue an exact equality. In the PCP package used [134] the ADMM iterations, for an original data set **X**, terminate when the constraint, defined by:-

$$\frac{\|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F}{\|\mathbf{X}\|_F} \qquad (8.5)$$

reaches a user defined tolerance.

There are, however, practical limits on how small this tolerance should be made. In addition to computational precision, the resolution of the pixels in an original image limit the accuracy. For example, each sample in the sectorial data, used here, is recorded as an integer (in the range 0 to 255) so that quantisation can induce an error of 0.5 to any pixel value. It is suggested [133] that in these cases a more convenient way to terminate the iterations is via the l_∞ norm. The constraint therefore becomes:-

$$\|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_\infty \quad \leq 0.5 \qquad (8.6)$$

Using this method the number of iterations for the full sector data was large and consequently time consuming, with later iterations producing very little reduction in the rank of L which begins to reach an asymptotic value. It was also found that the number of iterations defined by equation 8.6 was similar to that for equation 8.5 with a tolerance set to 0.0001. To assist with the comparisons the termination condition uses this tolerance with equation 8.5. In either case the number of iterations remains large (typically greater than 3000) with correspondingly long execution times (up to 2 hours). Methods of improving the performance of the robust PCA procedure is an area of research [138]. However the performance improvements do not appear to be significant so for this work there is little incentive to interfere with the provided algorithm. This is, however, a potential area for future research.

In its presented form the computational complexity of PCP limits its use to batch processing. That is, the low rank and sparse representations are derived from a fixed sequence of images over each unit of time, repeated computations of the sparse matrix, which contains direct evidence of anomalies, being required for anomaly detection. Although real time implementations of PCP are under review [139], [140] in this application it has potentially more direct and efficient use. This is by using the low rank representation as an outlier free training set for conventional PCA. Unlike surveillance applications where the background

may change slowly due to, for example, lighting conditions, the internal structure of an anomaly free weld will remain statistically constant.

In the previous investigations the sectorial data has been organised as projections of full images (sections 6.23 and 7.8) or constant A-scans (sections 6.24 and 7.9). Similarly TFM projections are of full images (section 6.26) or constant offsets (6.29). The same pattern of analysis applies in the following.

8.8   Sectorial data sets

PCP is now applied to the sectorial data sets. In this case the wedged probe is in direct contact with the test piece and the acquisition timing gates out the front wall. Each A-scan is recorded at the same gain and has equal significance so that beyond mean centring no attempt is made at data standardisation.

8.8.1   Full sector projections

For the sectorial data the training set is the same as that used in the trimming example (section 7.8). Once again each training set column is a concatenation of all columns from the respective sector. The 1816 sectors are derived from the two test blocks. As each sector contains 71 A-scans of 1544 samples the training set matrix is 109634×1816. After mean centring this is decomposed into its low rank and sparse representation. Conventional PCA is now a downstream operation in which the low rank eigenvectors represent the subspace of the anomalous free condition.

Equation 8.4 gives a default value for λ of 0.003. For the corresponding low rank matrix (L) the scree chart, figure 8.5.b, indicates that the ordered eigenvalues have a very steep descent. This is a characteristic of a data set where the observations are much smaller than the dimensions but here the situation is further exacerbated by a reduction in rank from over 1600 to 11. The spiked pattern, whereby the first few eigenvalues appear disproportionately large, yields the Kaiser stopping rule ineffective. Studies of these spiked eigenvalue structures are reported [141] but in this case it is both practical and simple to use all principal components.

a) Low rank - PCs 1-11, $\lambda = 0.003$ b) Trimmed - PCs 1:39
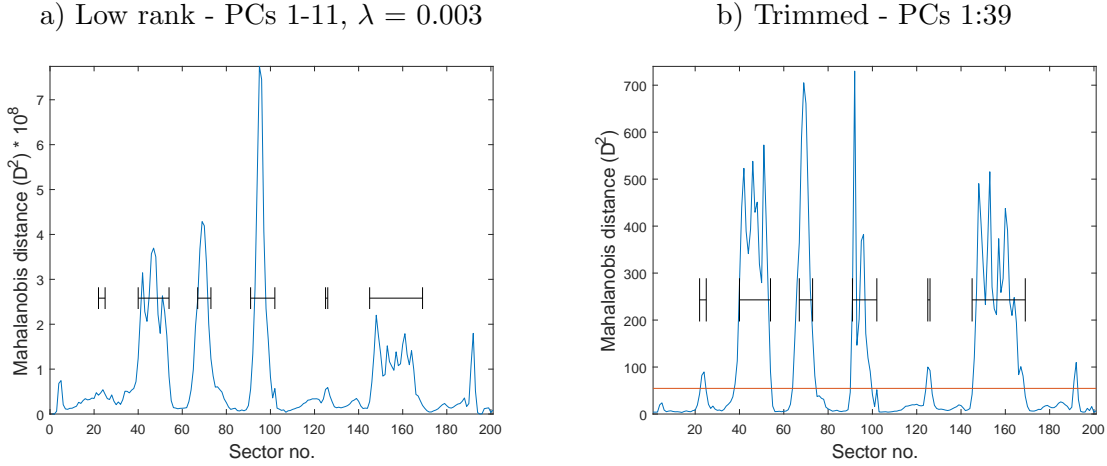
Figure 8.3: Comparison of low rank anomaly detection vs trimming

8.8.2 Projections using default λ

Figure 8.3 illustrates the projection of test piece TB1PA1 against the low rank sub-space for all principal components (1 to 11). For convenience the same result from the trimming example (section 7.8) is reproduced for comparison. Both plots have a similar shape although the trimmed example gives a more pronounced identification of the two small anomalies centred at sectors 23 and 122. Of more significance is the scale of the Mahalanobis distance, the maximum value extending to almost 8.10^8. A threshold based on the $\chi^2_{(pc,0.95)}$ (less than 20) for 11 PCs would result in an anomalous classification for practically every observation.

a) ROC curve   b) ROC threshold
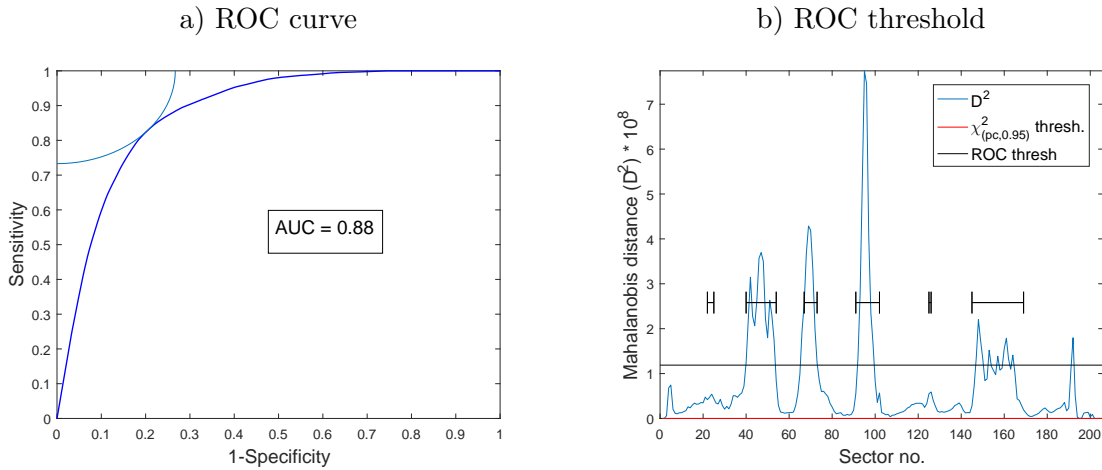
Figure 8.4: ROC curve and ROC threshold for TB1PA1 ($\lambda = 0.003$)

Further evidence of the unsuitability of this low rank subspace is indicated by ROC analysis. Not only does the curve (figure 8.4.a) have a poor AUC but the sensitivity and specificity figures (0.83 and 0.79 respectively) are also poor, as suggested by the ROC threshold level, figure 8.4.b. Projections of all other test sets (not illustrated) against the same subspace reveal similar results.

In the original work of Candès [107] the focus is on recovering a true low rank L and sparse S. A study by Paffenroth et al. [113] 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. Furthermore, the reduction in rank from over 1600 to 11 appears excessive for anomaly detection. Paffenroth et al. further suggest, and demonstrate, 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 they dealt with a single application and had 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. These can then be compared using a ROC curve as in section 7.8.4.

8.8.3 Effect of λ on rank reduction

The family of curves, figure 8.5.a, provides an overview of rank reduction with each iteration of the PCP algorithm for different values of λ. Overall these demonstrate that applying greater weighting to S does reduce the final rank of L. Of more significance is the fact that the initial relatively large reduction in rank is not maintained as the number of iterations increases. For the lower values of λ (approximately 0.003 to 0.008) it was rarely possible to satisfy the constraint (equation 8.6). This was either due to an excessive number of iterations (each taking approximately 25 seconds) or eventual instability as the error reduces. For this reason the number of iterations illustrated is limited to 100 with the error constraint being defined by equation 8.5 rather than equation 8.6. In all cases the errors converge giving some compatibility for comparisons. In the case of higher values of λ (0.015 and above) there are two points of significance. Firstly, the constraint defined by equation 8.6 is always reached, typically in less than 15 iterations, and secondly, the rank does not undergo such severe reduction. This is evident for the rank reduction curve

215

corresponding to $\lambda = 0.03$ (figure 8.5.a) which reaches the constraint after 8 iterations.

a) Rank reduction with iterations
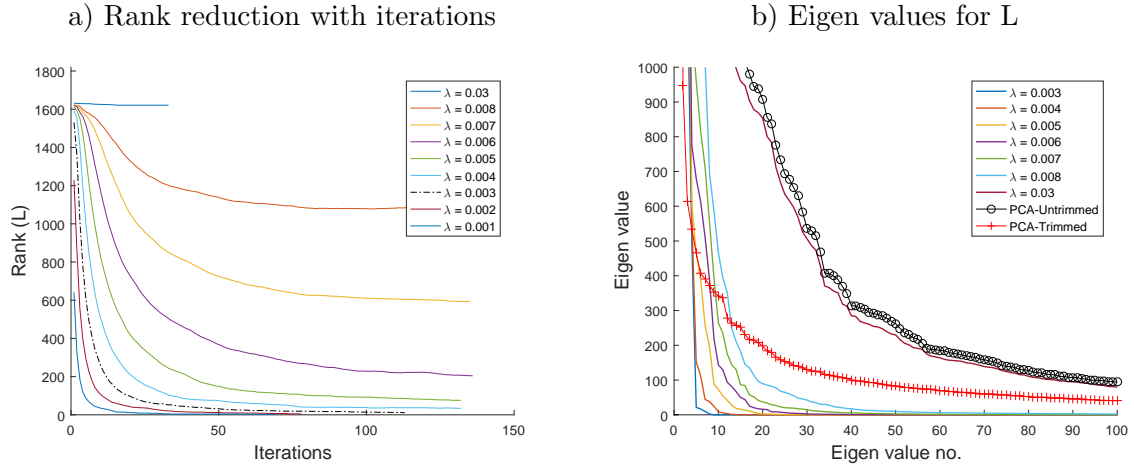
b) Eigen values for L

Figure 8.5: Eigenvalues and rank for L

8.8.4 Constraint error

Before excluding the lower values of λ from further discussion some attention is first given to the corresponding error curves, figure 8.6.a. Of interest is the fact that these do not reduce monotonically. After as little as three iterations the curves for λ's 0.003 to 0.008 have a local minimum around the third iteration (for 0.003 this is more representative as a point of inflection). The suggestion is that the local minimum, achieved very quickly, is a result of the major outliers being absorbed in a manner indicated in the introduction, section 8.1. If this is the case then the effect of training set outliers may be diminished in as few as 3 iterations, representing a significant saving in computational effort. Figures 8.6.b, c and d overview the results of a ROC analysis after projecting all test data sets after 2, 3 and 5 iterations. In each case a sequence of λ's are used (these are in the range 0.001 to 0.009 and 0.01 to 0.05, 15 in total). The curves do show that the corresponding ROC curves have peaks corresponding to λs in the expected range. However they never exceed 0.87 and are too low for accurate discrimination. For the higher values of λ increasing the iterations beyond 30 or more does result in improvements which are discussed next.

8.8.5 PCA using higher rank L

After establishing that a small number of iterations is not satisfactory and that acceptable error reduction is not possible with smaller values of λ, attention turns to investigating

216
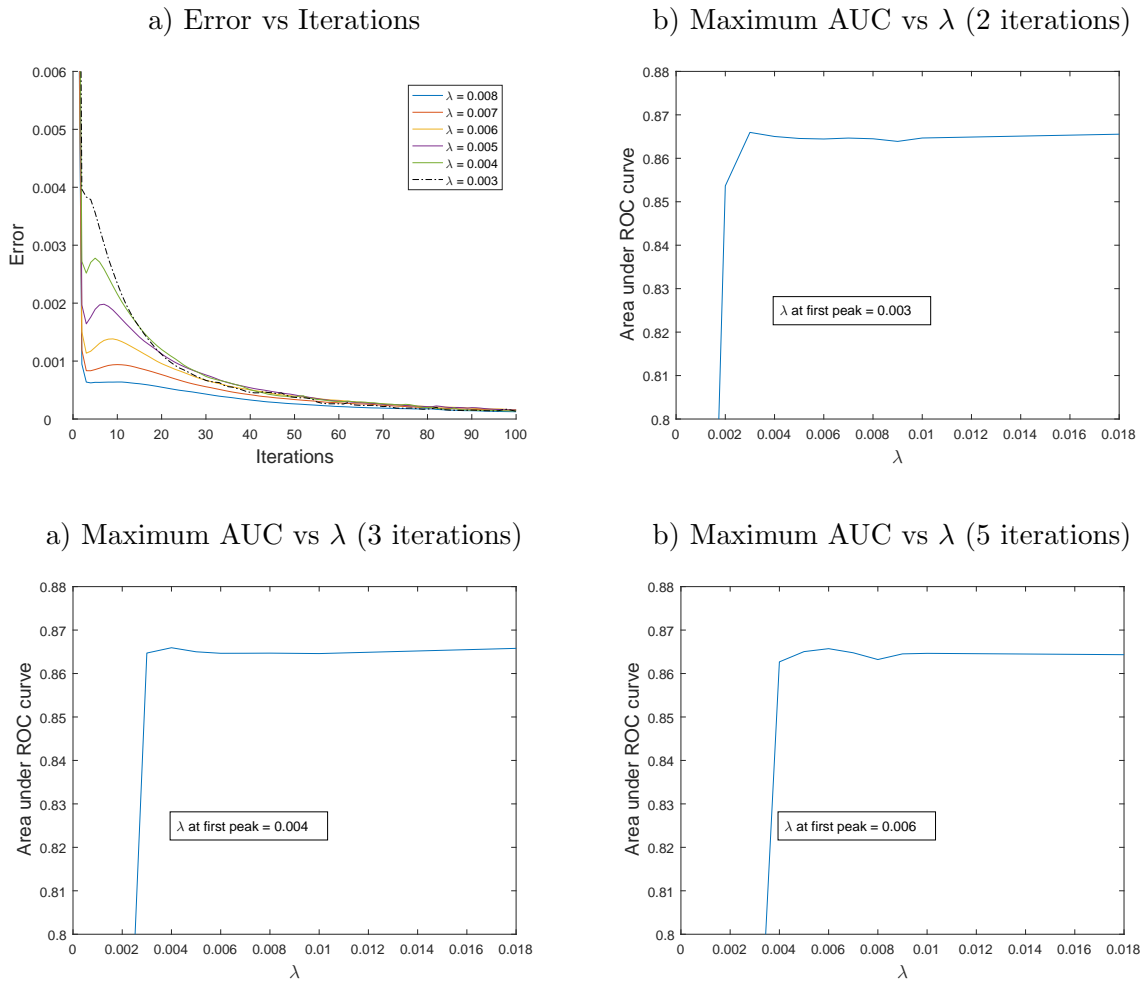
Figure 8.6: Investigation of first local minimum in error

higher values. Although this, in turn, has the effect of placing more emphasis on S the error constraint limits the rank reduction of L.

Once again the results are summarised by a set of ROC curves. After a number of experimental iterations it was found that increasing λ does improve the classification and that little further improvement is achieved after a value of around 0.02. Results are summarised by the family of ROC curves, figure 8.7.a and table 8.4.

With $\lambda = 0.01$ and above the first few eigenvalues remain spiked, but not to the same extent as for the lower values, figure 8.5.b. The Kaiser stopping rule now produces the values indicated in the table.
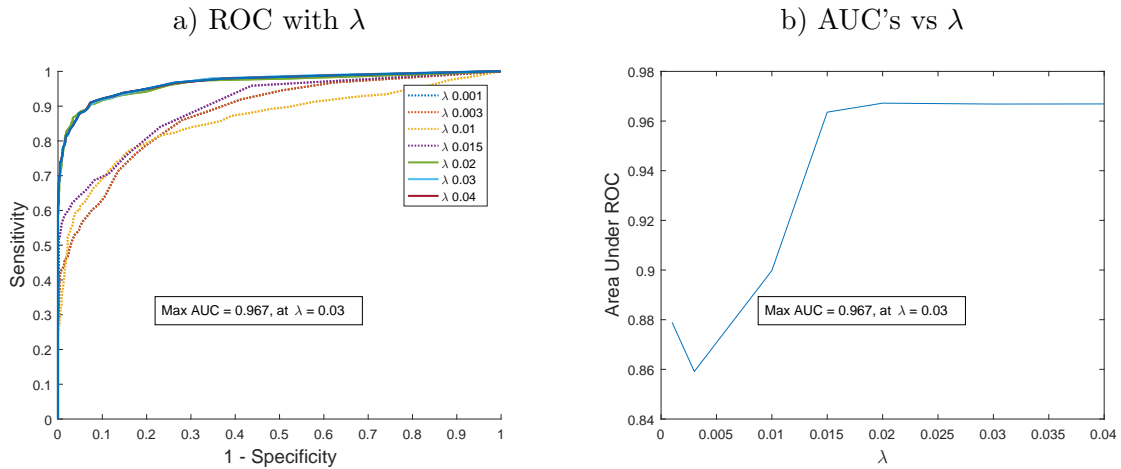
Figure 8.7: ROC comparisons with different λs

λ	Rank (L)	PCs	Thresholds		AUC	Acc(%)	Sens.	Spec.	F1
			$\chi^2_{(pc,0.95)}$	ROC					
0.001	2	2	5.99	1.44×10^6	0.87	78.4	0.79	0.80	0.72
0.003	11	11	19.67	36×10^6	0.85	80.6	0.77	0.84	0.74
0.01	1441	12	21.03	2.20	0.90	78.1	0.84	0.77	0.73
0.015	1580	32	46.19	10.01	0.96	90.6	0.90	0.93	0.88
0.02	1621	37	52.19	9.35	0.96	90.8	0.90	0.93	0.88
0.03	1626	39	54.57	9.02	0.96	90.9	0.91	0.93	0.88
0.04	1629	39	54.57	8.91	0.96	90.8	0.91	0.93	0.88

Table 8.4: Classification statistics for selected values of λ

8.8.6 Projection examples

In the following examples the low rank matrix is produced using $\lambda = 0.03$ and the number of principal components used is from the Kaiser stopping rule. Figure 8.8.a shows the result of projecting TB1PA1 against the low rank sub-space. Only the larger anomalies are captured by the $\chi^2_{(pc,0.95)}$ confidence level. Visually there is little evidence of the two smaller anomalies. Although the ROC threshold is lower it remains too large for a distinctive capture of the two smaller anomalies.

The spiked nature of the eigenvalue scree chart gives some concern regarding the number of principal components. To compensate, the number of PCs to use is doubled. The resulting projection, for $\lambda = 0.03$, is illustrated in figure 8.8.b. Now all anomalies have some visual presence but the raised confidence level is too high to capture the small instances. The ROC analysis indicates that a more appropriate level is possible and table 8.5 lists the new
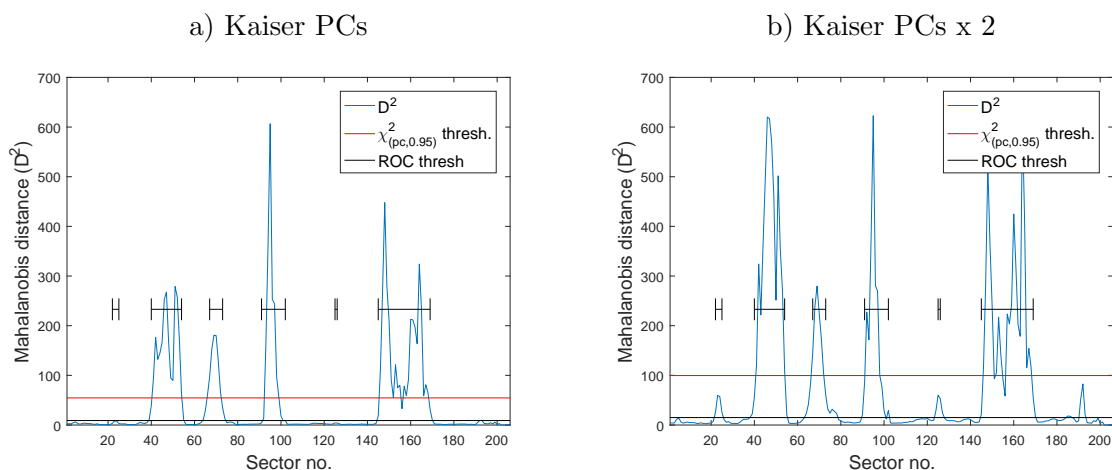
Figure 8.8: Example projections of TB1PA1

			Thresholds						
λ	Rank (L)	PCs	$\chi^2_{(pc,0.95)}$	ROC	AUC	Acc(%)	Sens.	Spec.	F1
0.001	2	2	5.99	1.44×10^6	0.87	78.4	0.79	0.80	0.72
0.003	11	11	19.67	36×10^6	0.85	80.6	0.77	0.84	0.74
0.01	1441	24	36.42	2.2	0.87	79.2	0.84	0.77	0.73
0.015	1580	64	83.68	29.9	0.97	91.2	0.92	0.93	0.89
0.02	1621	74	95.10	29.1	0.98	92.4	0.93	0.94	0.91
0.03	1626	78	99.62	27.84	0.98	92.1	0.94	0.94	0.91
0.04	1629	78	99.62	27.55	0.98	92.1	0.94	0.94	0.91

Table 8.5: Classification statistics for selected values of λ using $2\times$ Kaiser PCs

results for each test piece. For the chosen λ a comparison with the previous results does show some improvement.

8.8.7 Comparison with trimming

It is particularly interesting to note the similarity between the projections using PCP and those from trimmed data set (figures 8.8.b and 8.3.b. respectively). These are both on a similar scale and provide similar indications. They illustrate that both approaches have potential for detecting anomalies. In the case of PCP it was expedient to obtain overall classification results for the full set of eight test pieces. For this reason the comparison with the trimming case is the single row of average values. The full set of results is listed in table 8.6 where the PCP results are for $\lambda = 0.03$. Results from the trimming exercise are from the last row of tables 7.4, for the χ^2 confidence level, and 7.5, for the ROC analysis.

Although the ROC results are not practical they indicate the potential of the approach assuming an appropriate thresholding method can be found.

Method	PCs used	Threshold	AUC	Acc(%)	Sens	Spec.	F1
PCP	39	ROC	0.96	90.8	0.90	0.93	0.88
	78	ROC	0.98	92.1	0.94	0.94	0.91
Trimming	39	χ^2		93.6	0.91	0.93	0.89
	39	ROC	0.97	93.5	0.92	0.94	0.90

Table 8.6: Comparison of classification results between PCP and trimming

8.8.8 Review

Consistent with the work of Paffenroth et al. [113] (section 8.8.2) this section has demonstrated that a low rank matrix produced by RPCA has some potential for anomaly detection. However it is necessary, through experiment, to establish a value for λ. Care also needs to be given to the constraint criteria. Continual iterations to reduce the error reduces the rank far too much. In this particular case the maximum number of iterations were found to be best limited by equation 8.5. However two problems remain. First the spiked nature of the eigenvalues does not facilitate selection of the number of PCs to use for the projections, the indications being that the Kaiser stopping rule results in a number too small. Doubling the number of Kaiser PCs produced more acceptable results with evidence of all anomalies appearing in the projection. The second problem relates to selection of a threshold. The $\chi^2_{(pc,0.95)}$ value remains too high.

It may be possible to overcome these problems following further investigation. However the number of dimensions makes considerable demands on computing effort and the process is particularly time consuming. According to the comparison in table 8.6 classification using PCP is potentially slightly better than that for trimming. With PCP, however, the $\chi^2_{(pc,0.95)}$ confidence level is too high resulting in very low specificities and sensitivities. This does not present a problem in the data trimming case where the $\chi^2_{(pc,0.95)}$ classification results compare favourably with those suggested from ROC analysis.

8.9 Constant A-scan projections

An approach similar to that presented in section 7.9 is now followed and analysis of the results follows the same pattern. The main difference is in establishing the training set.

Previously this was done by trimming the original training set to remove outliers. Here the low rank representation of the training set is produced by PCP.

For review purposes the training set is a set of constant angle A-scans taken from all sectors. Each A-scan is a column vector with each sample point representing a dimension or variable, the number of observations being represented by the number of sectors in the data set. With the present set up each A-scan contains 1544 samples and a significant advantage of this approach, over full sector projections, is a considerable reduction in the number of variables. A further advantage is that location of an anomaly is by sectors and A-scans rather than by sector alone. The disadvantage is that a training set is required for each A-scan angle (71 in the present case).

From the total data available (1816 sectors) seventy one training sets, each corresponding to its respective A-scan angle, are established. Each is mean centred before processing to produce a low rank training set using the previously described PCP algorithm.

Section 8.8.5 demonstrated that the default value of λ places too much emphasis on the S matrix. It was also demonstrated that there is no point in executing more iterations of the PCP algorithm beyond the condition specified by equation 8.6. Before proceeding it is necessary to ascertain if both of these conditions apply to this case.

8.9.1   Evaluation of λ

Low rank training sets are first established using the default value of λ which in this case is approximately 0.023. After some initial investigation a tolerance level of 0.0001 was established for use with equation 8.5. This represents a point at which the error reaches an asymptotic value. Each A-scan from each test piece is projected using the respective low rank sub-space and the corresponding Mahalanobis distances are recorded. The image, figure 8.9.a, is created from the Mahalanobis distances of each A-scan across all sectors for TB1PA1 whilst in the case of figure 8.9.b each $\chi^2_{(0.95,PC)}$ confidence level has been subtracted. Neither result is helpful. For sector projections much improved results were obtained for higher values of λ. Following these findings a value of $10\times\ \lambda_{default}$ gave a considerable improvement. Figures 8.9.c and d represent a repeat of the previous test with $\lambda = 0.25$. As with full sector projections this value appears to represent an almost optimum value. Although arrived at by trial and error small changes around this value make very little difference to the results. All PCP results now presented use $\lambda = 0.25$.

For constant A-scan projections two look-up methods are available, namely sector only and

sector with A-scan. Consistent with the trimming examples (section 7.9) results are first presented based on sector only look-up followed by sector and A-scan look-up.
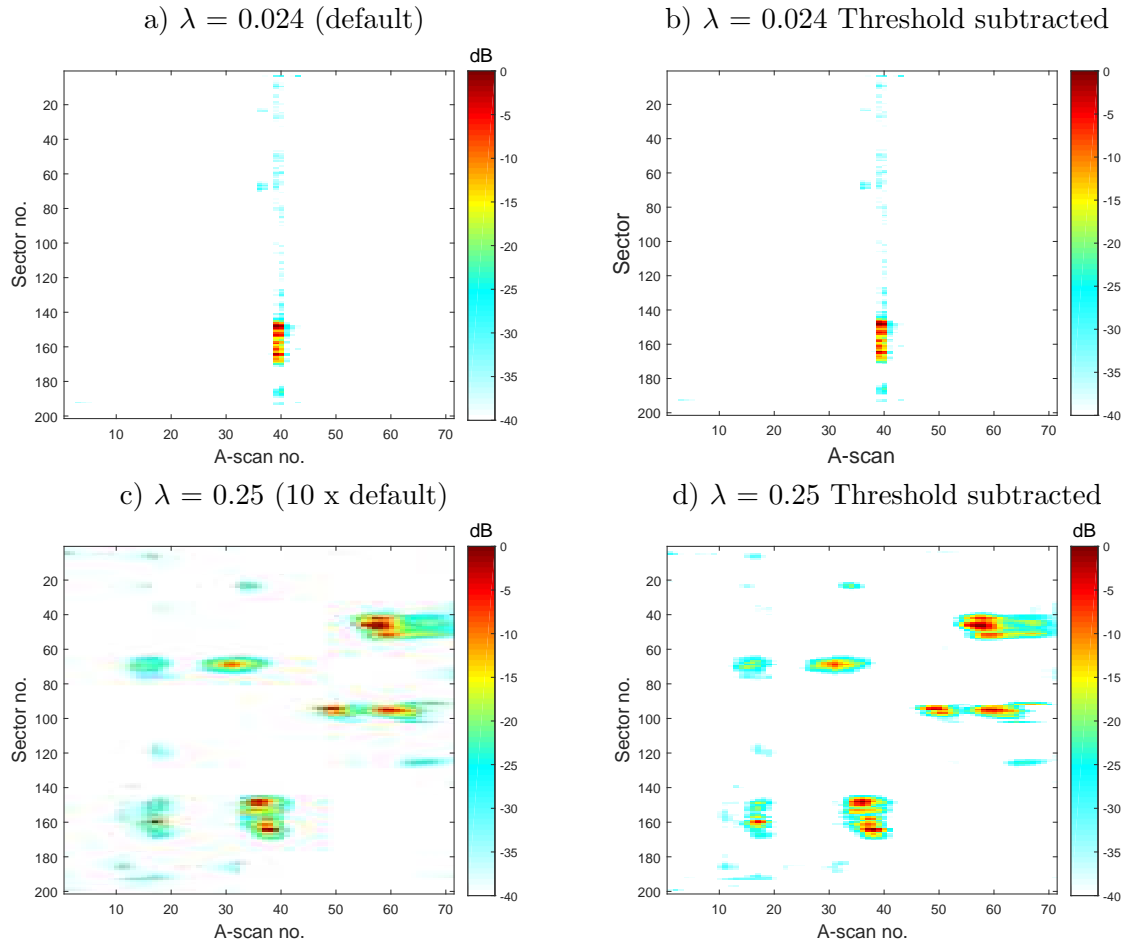
Figure 8.9: Constant A-scan projections using default λ and 10*default

8.9.2 Sector only look-up

To account for the different number of PCs associated with each A-scan the respective $\chi^2_{(pc,0.95)}$ confidence value is subtracted before analysis proceeds further. Now any value greater than zero contains a potential anomaly. For sector only look up the maximum D^2 distance across all sectors is extracted. Figure 8.10 illustrates the result of this for TB1PA1. These projections are remarkably similar in shape to those using trimmed data (figures 7.5). One significant difference is that now far fewer projections exceed the χ^2 confidence level,

reducing the number of false positives. This suggests that the subspace, using PCP, is possibly less biased and more accurate than that determined through trimming.

a) Original projections   b) Projections with $\chi^2_{(pc,0.95)}$ subtracted
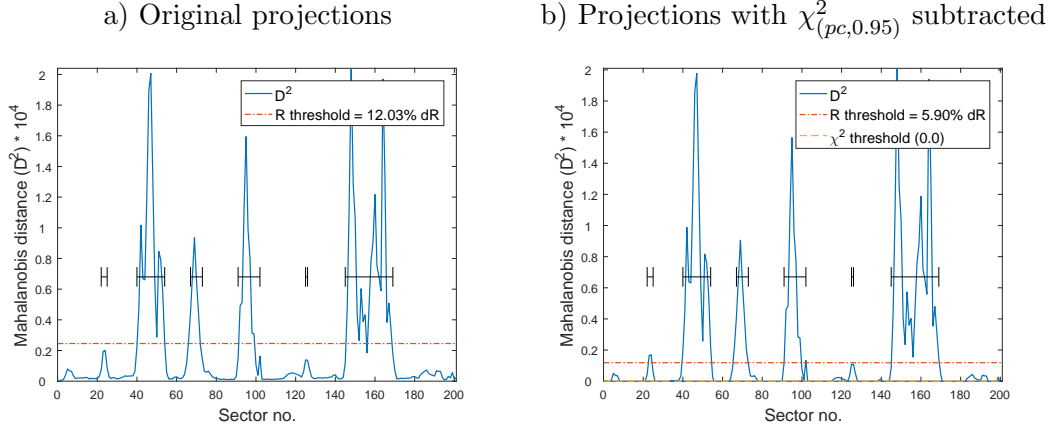
Figure 8.10: Example projections (TB1PA1) using low rank subspace $\lambda = 0.25$

Table 8.7 compares the classification statistics for sector only look-up using PCP to those from trimming. The trimming results are duplicated from table 7.11. Both sets of results use the χ^2 confidence level alone. The PCP results represent a considerable improvement, although the specificity figures continue to indicate that the approach produces too many false positives for practical use.

Results of a ROC analysis, table 8.8, indicate the potential of this approach. Whilst the ROC analysis shows only the potential of the classification without a practical method of achieving this the results of PCP are more accurate. These remarks, however, only apply to situations were the intention is to identify anomalous sectors numerically. For a visual interpretation either figures 7.4.c from trimming or 8.10 from PCP provide indicators for further manual investigation.

223

Test Piece	Trimming $\chi^2_{(pc,0.95)}$				PCP $\chi^2_{(pc,0.95)}$			
	Acc.(%)	sens.	spec.	F1	Acc.(%)	sens.	spec	F1
TB1PA1	40.8	1.0	0.13	0.52	76.6	1.00	0.65	0.73
TB1PA2	43.8	1.0	0.05	0.59	70.2	1.00	0.47	0.73
TB1PA3	46.9	1.0	0.21	0.55	74.6	1.00	0.62	0.72
TB1PA4	42.4	1.0	0.16	0.52	77.0	1.00	0.67	0.73
TB2PA1	36.8	1.0	0.10	0.49	73.0	1.00	0.62	0.69
TB2PA2	37.8	1.0	0.07	0.52	70.6	0.99	0.56	0.69
TB2PA3	49.8	1.0	0.25	0.57	84.9	1.00	0.77	0.81
TB2PA4	40.8	1.0	0.19	0.47	70.0	1.00	0.60	0.64
Averages	42.3	1.0	0.14	0.53	74.6	1.00	0.62	0.72

Table 8.7: Sector only look-up classification trimmed (table 7.11) vs. PCP

Test piece	$\chi^2_{(pc,0.95)}$				ROC					
	Acc(%)	Sens.	Spec.	F1	R_{th}(% dR.)	AUC	Acc(%)	Sens.	Spec.	F1
TB1PA1	76.6	1.00	0.65	0.73	5.90	0.98	93.9	0.95	0.93	0.91
TB1PA2	70.2	1.00	0.47	0.73	5.37	0.97	92.1	0.95	0.90	0.91
TB1PA3	74.6	1.00	0.62	0.72	2.81	0.98	93.7	0.95	0.93	0.91
TB1PA4	77.0	1.00	0.67	0.73	2.56	0.98	93.7	0.95	0.93	0.90
TB2PA1	73.0	1.00	0.62	0.69	5.56	0.98	93.4	0.96	0.92	0.90
TB2PA2	70.6	0.99	0.56	0.69	1.0	0.97	92.4	0.93	0.92	0.89
TB2PA3	84.9	1.00	0.77	0.81	1.7	0.99	95.8	0.97	0.95	0.94
TB2PA4	70.0	1.00	0.60	0.64	5.08	0.99	95.1	0.97	0.95	0.91
Averages	74.6	1.00	0.62	0.72	3.75	0.98	93.8	0.95	0.93	0.91

Table 8.8: Sector only look-up ROC analysis

8.9.3 Sector and A-scan (full image) look-up

Figure 8.11.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. Visually the indications are that all anomalies are captured with very few false positives. Figure 8.11.b is the corresponding image (on a dB scale) for projections without zero adjustment. Using areas specified by the rectangles, table 8.9 lists the SNR of each anomaly. For convenience the table also lists the SNRs of previous investigations.

Overall the SNRs are an improvement over those for the raw A-scans suggesting a potential for improved discrimination. Although not quite as good as those from the trimming exercise they are, nevertheless, in keeping. This and the relatively clean 'floor' of figure 8.11.a indicates that discrimination is likely to be an improvement over trimming.

Evaluation of full image look-up proceeds in the same manner as for trimming. First the

a) Normalise projections above confidence limit   b) Full dB image (no threshold)
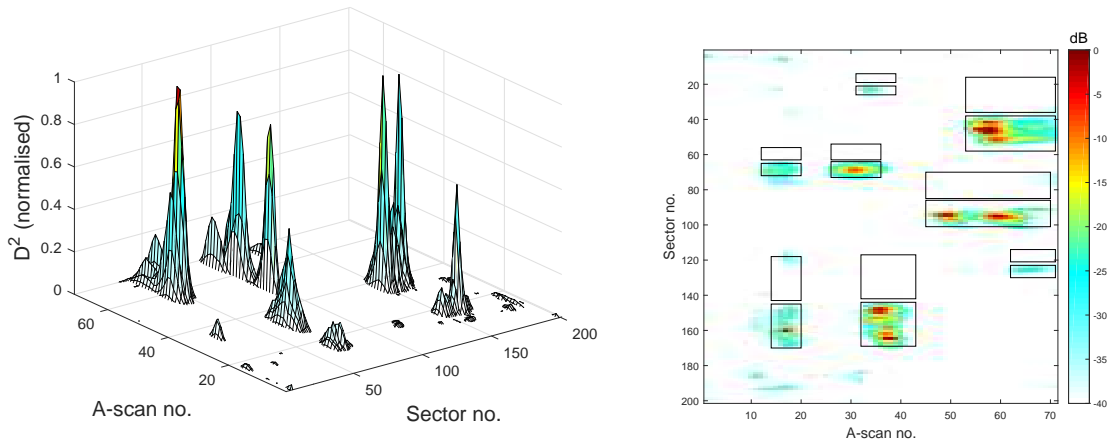
Figure 8.11: TB1PA1 projections ($\lambda = 0.25$)

Fault Instance	A-scan Peaks	SNRs (dB) D^2	
		Trimmed	PCP
Pore 1	23.67	43.1	28.1
Crack	31.25	51.0	44.8
Porosity I1	24.22	52.0	41.4
Inclusion 1	29.58	50.2	42.2
Pore 2	18.78	33.0	21.1
Inclusion 2 I1	35.82	61.2	49.6
Porosity I2	18.54	36.9	28.7
Inclusion2 I2	25.67	39.8	37.5

Table 8.9: SNR comparisons (trimmed vs PCP)

image created from the full set of projections (e.g figure 8.11.b) is converted to binary, using the χ^2 confidence level as a threshold. It is then compared with the ground truth. The results for each test piece are given in table 8.10 under the PCP $\chi^2_{(pc,0.95)}$ columns. These once again represent an improvement over the equivalent case for trimming, duplicated under the Trimming $\chi^2_{(pc,0.95)}$ columns from table 7.13, and are better than those for sector only look up (table 8.8).

With sensitivities of every test piece having a value of 1 the indication is that, as a threshold, the χ^2 confidence level is too low. The results of a ROC analysis (table 8.8) indicate that with an increase in threshold near perfect discrimination is possible. At the expense of a 1% drop in sensitivity the minimum specificity is now 0.98. The F1 scores indicate a reliable ROC analysis and reduction in false positives.

Test Piece	Trimming $\chi^2_{(pc,0.95)}$				PCP $\chi^2_{(pc,0.95)}$			
	Acc.(%)	sens.	spec.	F1	Acc.(%)	sens.	spec	F1
TB1PA1	83.9	1.00	0.83	0.33	95.7	1.00	0.96	0.65
TB1PA2	69.5	0.99	0.93	0.22	91.2	1.00	0.91	0.51
TB1PA3	84.7	0.99	0.84	0.35	94.3	1.00	0.94	0.59
TB1PA4	86.0	0.99	0.86	0.35	94.2	1.00	0.94	0.56
TB2PA1	81.3	0.92	0.81	0.28	94.4	1.00	0.94	0.58
TB2PA2	81.3	0.99	0.80	0.37	93.9	0.99	0.94	0.64
TB2PA3	89.0	0.94	0.90	0.36	96.5	1.00	0.96	0.66
TB2PA4	87.4	0.91	0.87	0.31	95.8	1.00	0.96	0.60
Averages	82.9	0.97	0.86	0.32	94.5	1.00	0.95	0.60

Table 8.10: Classification trimmed (table 7.13) vs. PCP

TestPiece	PCP $\chi^2_{(pc,0.95)}$				ROC				
	Acc. (%)	Sens.	Spec.	F1	AUC	Acc(%)	Sens.	Spec.	F1
TB1PA1	95.7	1.00	0.96	0.65	0.99	98.5	0.99	0.99	0.84
TB1PA2	91.2	1.00	0.91	0.51	0.99	97.6	0.99	0.98	0.79
TB1PA3	94.3	1.00	0.94	0.59	0.99	98.4	0.99	0.98	0.84
TB1PA4	94.2	1.00	0.94	0.56	0.99	98.2	0.99	0.98	0.81
TB2PA1	94.4	1.00	0.94	0.58	0.99	98.0	0.99	0.98	0.79
TB2PA2	93.9	1.00	0.94	0.64	0.99	98.3	0.99	0.98	0.87
TB2PA3	96.5	1.00	0.96	0.66	0.99	99.2	0.99	0.99	0.86
TB2PA4	95.8	1.00	0.96	0.60	1.0	99.3	0.99	0.99	0.89
Averages	94.5	1.00	0.95	0.60	0.99	98.5	0.99	0.98	0.84

Table 8.11: PCP vs ROC analysis

Without a method of obtaining a threshold automatically the ROC results are only theoretically possible. Table 8.12 compares the threshold values (as a percentage of the total dynamic range) for the three thresholding methods against those from the ROC analysis.

Once again the KI method produces the closest match, the Otsu and ME methods being far too high. Results using KI thresholds did not approach those of the ROC analysis. They are therefore compared with the originals using the χ^2 confidence level (table 8.13). In all cases the KI threshold produces a sensitivity of 1 and a specificity lower than that for the χ^2 confidence level. That is, the KI threshold is too low and, more importantly, provides no improvement over the χ^2 confidence level.

Test Piece	R_{th}	KI_{th}	O_{th}	ME_{th}
TB1PA1	3.19	0.92	17.68	11.77
TB1PA2	2.30	1.13	16.77	10.98
TB1PA3	1.20	0.33	11.89	9.80
TB1PA4	1.55	0.34	22.40	8.63
TB2PA1	2.82	1.04	24.38	16.47
TB2PA2	0.45	0.15	21.63	6.28
TB2PA3	1.15	0.24	17.40	7.45
TB2PA4	2.17	0.40	12.73	12.55

Table 8.12: Projection thresholds as a % of total dynamic range

TestPiece	PCP $\chi^2_{(pc,0.95)}$				KI			
	Acc(%)	Sens.	Spec.	F1	Acc. (%)	Sens.	Spec.	F1
TB1PA1	95.7	1.00	0.96	0.65	91.6	1.0	0.91	0.49
TB1PA2	91.2	1.0	0.91	0.51	92.3	1.0	0.92	0.53
TB1PA3	94.3	1.0	0.94	0.59	90.2	1.0	0.9	0.46
TB1PA4	94.2	1.0	0.94	0.56	88.3	1.0	0.88	0.39
TB2PA1	94.4	1.0	0.94	0.58	92.1	1.0	0.92	0.50
TB2PA2	93.9	1.0	0.94	0.64	90.4	1.0	0.90	0.54
TB2PA3	96.5	1.0	0.96	0.66	91.8	1.0	0.92	0.44
TB2PA4	95.8	1.0	0.96	0.60	89.9	1.0	0.90	0.38
Averages	94.5	1.0	0.95	0.60	90.8	1.0	0.91	0.47

Table 8.13: PCP vs PCP with additional KI thresholding

Figure 8.12 illustrates the outcome of thresholding the projections for TB1PA1 using each method. The close similarity between the ground truth and that using the ROC threshold is particularly noticeable. The result for this test piece and all others is confirmed by table 8.11. Although the ROC threshold can only be determined in the presence of a ground truth these results indicate that PCP provides a considerable improvement over trimming. Unfortunately each method of thresholding is, in these cases, unsatisfactory.

An earlier discussion on thresholding (section 7.10.1.5) demonstrated that because the available TFM images are dominated by background their histograms are unimodal. The Otsu and ME methods, which rely on a bimodal histogram, are not suitable for thresholding such images. For sectorial data, earlier work using peak A-scan values (see for example
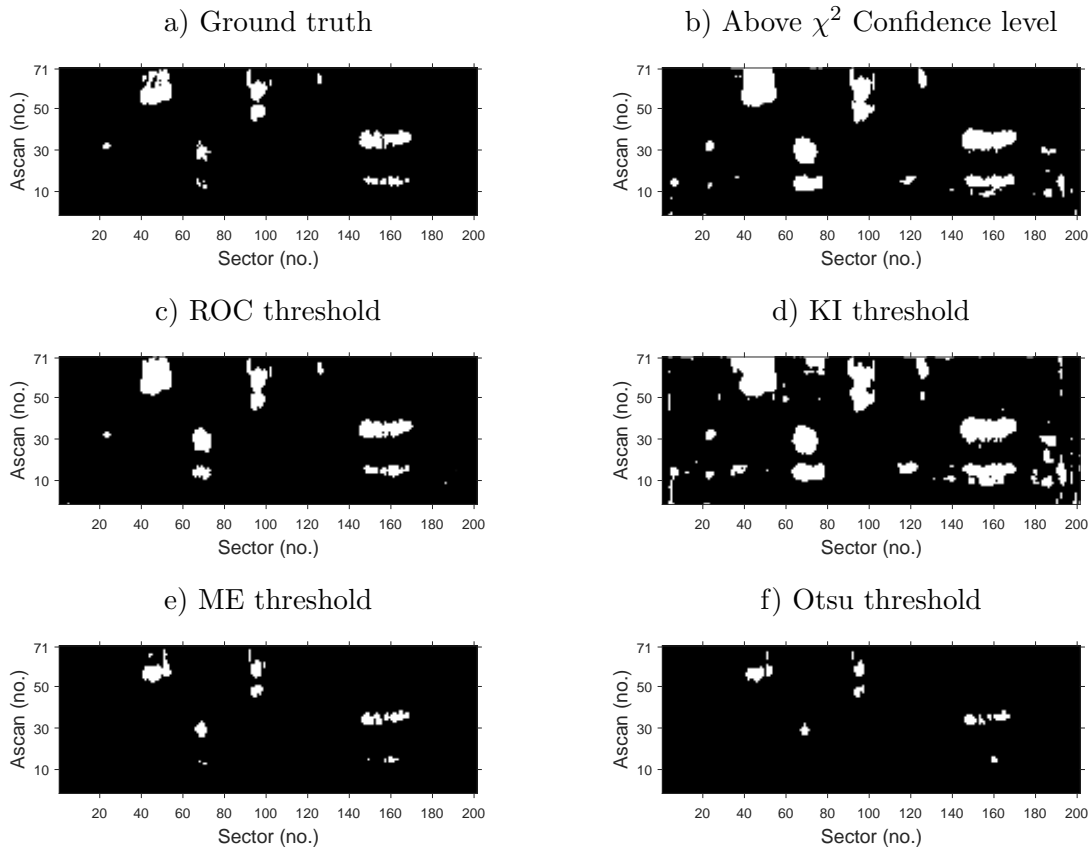
Figure 8.12: Threshold comparisons

figure 5.5), demonstrates that both the ME and Otsu thresholds produce results comparable to KI thresholding. With PCP's D^2 projections this is not the case. Examination of a typical histogram of these D^2 projections, figure 8.13 (restricted to distances ≤ 4000), indicates that, unlike the peak A-scan images, background dominates the histogram so that both the Otsu and ME threshold methods fail. Although the KI is a reasonable value for the threshold it is lower than that of the ROC.

If, following PCP, such histograms are typical then an alternative method of thresholding may be necessary. However providing the specificities are acceptable a solution may be to simply use the χ^2 confidence level alone. For example figure 8.12 indicates a relatively small number of false positives. These are unlikely to be too much of a distraction. For automation figure 8.14.a illustrates the result of blob detection. In this case all anomalies have been captured. There are an additional 14 regions of false positives. Based on size many of these can be rejected as noise. Blobs contained in other, larger, regions may need

further investigation. For comparison the result of blob detection after trimming (from figure 7.8) is also illustrated.
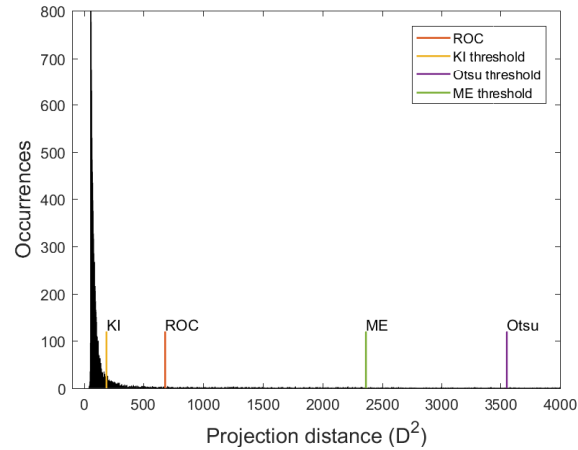

Figure 8.13: Histogram of typical projections with thresholds ($D_{max}^2 > 20000$)

a) PCP above χ^2 confidence level
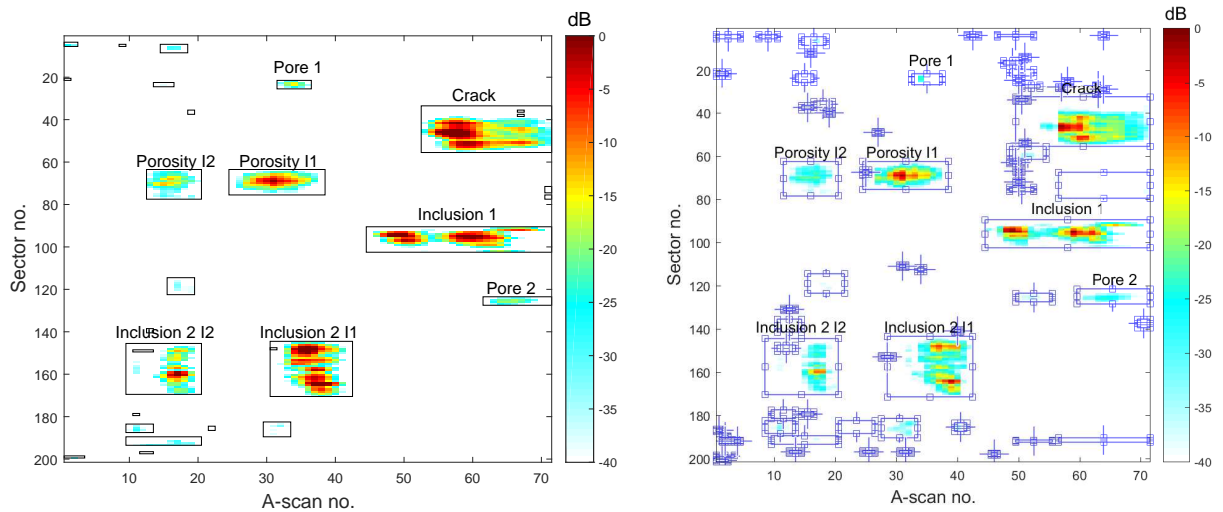
b) Trimming above KI threshold (fig 7.8)

Figure 8.14: Blob detection comparisons (TB1PA1)

8.9.4 Comments

For both full sector and constant A-scan orientations the tuning parameter, λ, as defined by equation 8.4 (and recommended by Candès et al. [107]) did not produce satisfactory results.

The fact that the resulting singular values of the corresponding low rank matrix were 'spiky' is an indication that the data set did not meet pre-requisite conditions for method. These and other pre-requisites are also noted in the reference paper [107]. After some trial and error using the sector only orientation a value of 10 times that given by equation 8.4 was found to provide much improved results. The same approach (i.e multiplication by 10) was also found to produce considerably improved results for the constant A-scan orientation. In fact it was also found that slight variations around this modified value have little effect on the results. With only a single data set it is not possible to verify whether this multiplication by 10 is specific to this data set or has more general use.

Unlike the sector only orientation, which did not produce results significantly better than those from trimming, the constant A-scan orientation gave significant improvements, particularly for sector and A-scan (full image) look-up. With the exception of the value of λ the process was mechanical. That is, the number of PCs to use is that determined by the Kaiser stopping rule. ROC analysis demonstrates that for all test pieces a threshold exists that produces a considerable improvement in discrimination. Although none of the selected thresholding methods is able to locate this value the χ^2 confidence limit itself provides a more accurate classification than that from trimming. This is evident by the results listed in table 8.10 with figure 8.14 providing a more emphatic visual comparison. It is emphasised that for PCP, blob detection is based on projections above the χ^2 confidence limit only, whereas in the case of trimming, blob detection is after KI thresholding.

8.10 TFM data sets

This section starts by presenting results for full image projections followed by those for the constant offset orientation. In each case PCP is used to recover a low rank subspace considered to be representative of the anomaly free background. Initially λ is set to its default value as specified by equation 8.4. The results are from the projection of observations using the eigenvectors of the low rank sub-space.

8.10.1 Full image projections

Full image data sets are created in the usual way, that is by first concatenating each column of an image into a single column vector. The complete data set is the collection of all column vectors for each image as a set of rows. Initial results are for projections of the data set using the low rank subspace. As usual the $\chi^2_{(pc,0.95)}$ limit is used as the reference

boundary. The distance measure used is the Mahalanobis distance but for standardised data this is equivalent to the Euclidean distance.

a) TFM1 ($\lambda = 0.0053$)

b) TFM2 ($\lambda = 0.0053$)

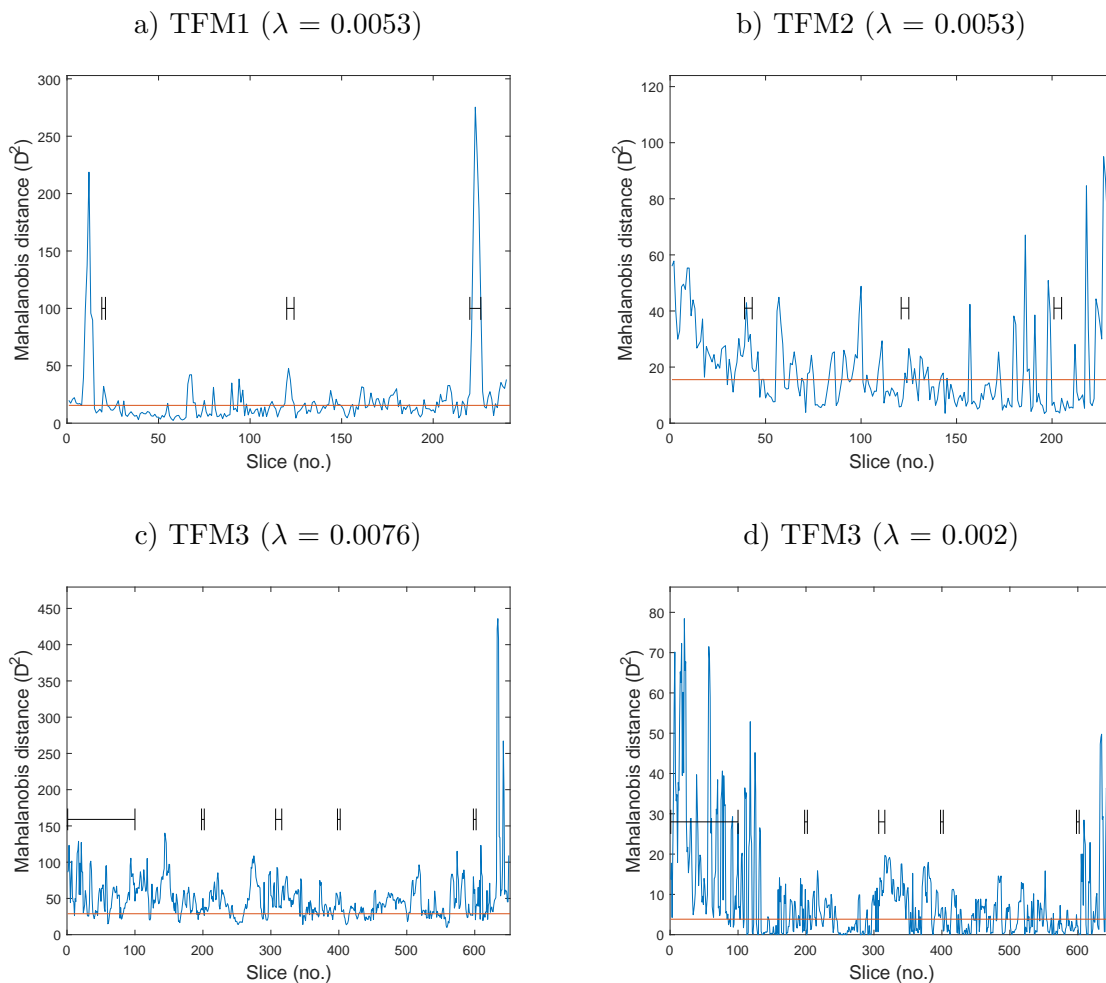c) TFM3 ($\lambda = 0.0076$)

d) TFM3 ($\lambda = 0.002$)

Figure 8.15: Full image projections for each TFM test set

Projections for each test piece are illustrated in figure 8.15. Consistent with the equivalent results for the trimmed data case (figure 7.9) these are not encouraging. Once again TFM1 suggests some indications of anomalies but equally there are numerous false positives. All images are included in the training sets and the anticipation is that the PCP procedure will suppress evidence of foreground features from the low rank representation.

Unlike the bottom drilled holes the SDH in TFM3 persists over a large number of sequential images. One consequence of this is that some registration of the SDH is likely to be that of background rather than foreground. By contrast the shorter duration bottom drilled holes more readily register as foreground. Although variations along the SDH are likely to be

231

collected in the S matrix a significant low rank representation will also be present. To test this a greater bias is placed on the sparse(S) matrix (hence less on L) by reducing λ. The effect of this is illustrated in figure 8.15.d. However this has done nothing to improve the detection of the remaining SBHs.

Trial and error experiments on each test piece using different values of λ proved not to produce satisfactory results. In fact none were significantly better than those illustrated in figure 8.15. The approach of full image projection was, therefore, abandoned in preference to the constant offset orientation which is discussed next.

8.10.2   Constant offset orientation

To review, the constant offset orientation uses all observations from each offset to create the respective low rank subspace. Each distance (D^2) now results from the projection of the respective column vector against its offset's low rank subspace. The net effect can be represented as a 3 dimensional surface (figures 8.16.a, c, and e) or a 2 dimensional image (figures 8.16.b, d, and f). Although it is relatively small each set of figures has a different reference level. The 3D surface plots are with respect to the $\chi^2_{(pc,0.95)}$ confidence level for each offset. Each projection above zero represents a potential anomaly. The two dimensional images are not adjusted. They represent the entire dynamic range of the projections and may be used for estimating the SNR of each anomaly.

Visually, anomalies in the 3D projections lack immediate clarity. Although infested by the same level of noise the 2D representation gives an improved overview. In all cases these are in contrast to the equivalent sectorial image, figure 8.9.a which also uses default value of λ. A conclusion might be that, for the TFM images, the default value of λ is a suitable choice. A comparison of the SNRs listed in table 8.16 with those from the trimming exercise, and repeated here, indicate that some improvement might be possible.

Investigations of the sectorial data revealed that λ has an effect on discrimination and SNRs. For both full sector and constant A-scan orientations, a considerable improvement was achieved by setting λ to 10 times the default value. It was also found that, for the full sector orientation, the number of PCs specified by the Kaiser stopping rule had to be doubled. Similar investigations need to be made here.

This section now continues with an investigation of slice only look-up. This will, however, be short. Previous investigations have indicated that for the constant offset (or constant A-scan) orientation the slice and offset (or sector and A-scan) look-up method provides the
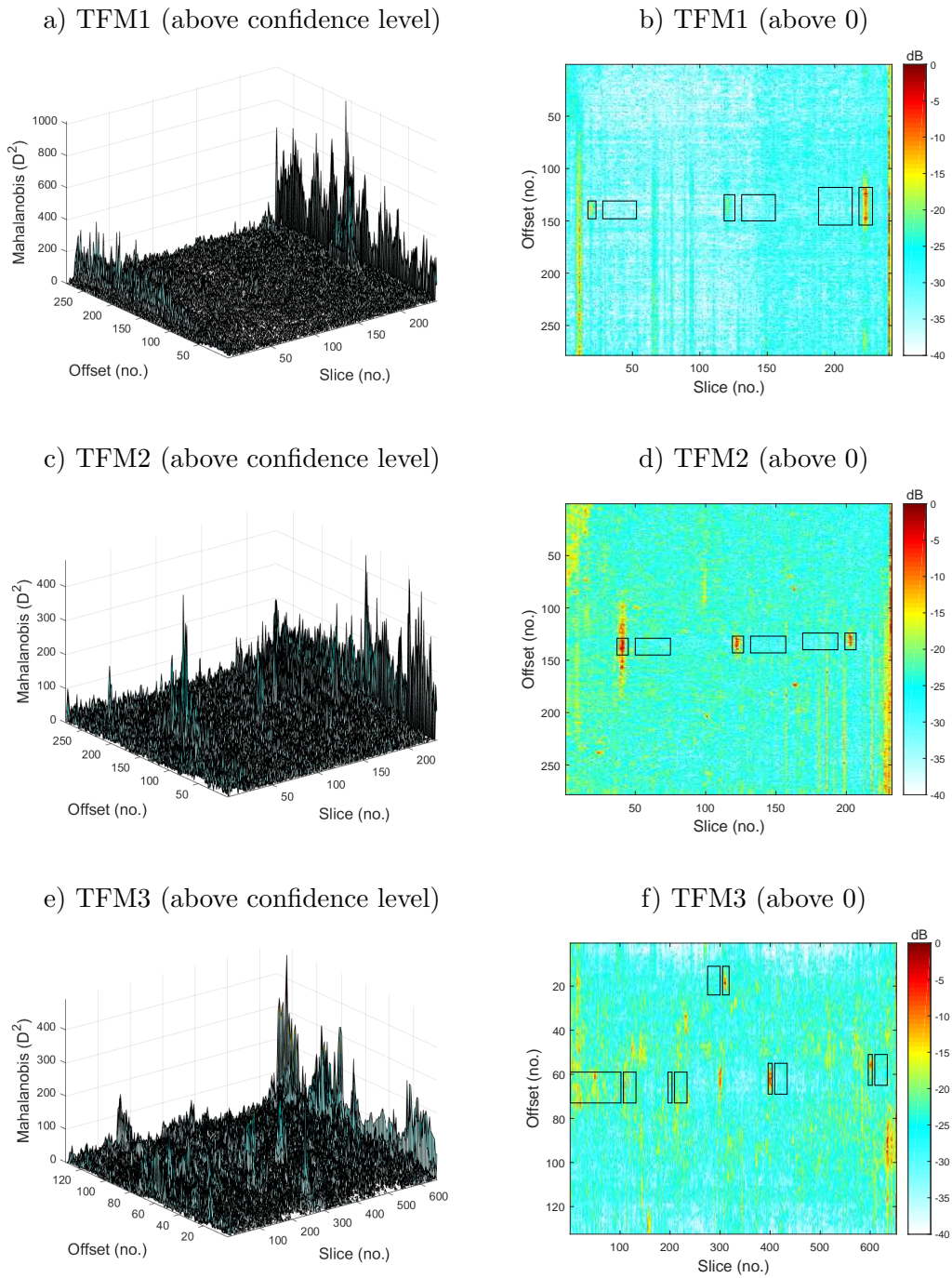
a) TFM1 (above confidence level)   b) TFM1 (above 0)

c) TFM2 (above confidence level)   d) TFM2 (above 0)

e) TFM3 (above confidence level)   f) TFM3 (above 0)

Figure 8.16: Constant offset projections with slice and offset look-up

233

most detail. Consequently, after making some short comments about slice only look-up, this will be the emphasis of the following.

8.10.3 Constant offset with slice only look-up

For slice only look-up the peak value of each slice (column) is plotted against slice number, figure 8.17. For each test piece the left hand figure is the peak of all offsets whilst the right hand figure is the projection value for a specific offset, chosen because it passes through most (or all) of the anomalies. In contrast to the full image projections these now indicate greater discrimination. This is particularly the case for the single offset case. However in practice, any offset containing an anomaly is not known in advance and it is the left hand figures, which use the peaks of all offset values, that are of more interest. Although this continues to give indications of true anomalies these are less well defined. In particular all projections are now above the base threshold with other spurious, large, projections adding further confusion. In the case of TFM3 there is little evidence of the SDH. This is likely to be for the same reason as discussed previously (section 8.10.1).

As all projections in figures 8.17.a, c, and e are above the threshold there are no true negatives and very many false positives. Under these circumstances there is little point in determining the confusion statistics. This is confirmed by the results of a ROC analysis listed in table 8.14 where the R_m column represents the ROC threshold as a proportion of that for the average $\chi^2_{(pc,0.95)}$ over all offsets. Using the Kaiser stopping rule an indication of the number of PCs for each offset is listed in table 8.15.

Test Piece	$\chi^2_{(pc,0.9)}$				ROC					
	Acc(%)	Sens.	Spec.	F1	R_m	AUC	Acc(%)	Sens.	Spec.	F1
TFM1	6.3	1.00	0.0	0.12	5.01	0.83	88.0	0.73	0.89	0.43
TFM2	6.5	1.00	0.0	0.12	5.28	0.90	84.5	0.87	0.84	0.42
TFM3	19.2	1.00	0.0	0.32	4.86	0.67	65.0	0.58	0.67	0.39

Table 8.14: ROC values for constant offset slice only look-up

As usual for the ROC analysis these results represent the best classification that might be achieved from the available data. It provides no assistance to determine the threshold.

Classification statistics for TFM3 are particularly poor. This is thought to be due to the fact that the SDH is not well defined. Although better figures are achieved for TFM1 and TFM2 all results remain relatively poor. The ROC analysis indicates that, for this look-up method, PCP has no better potential than that of trimming (table 7.27). In particular

a) TFM1 (max D^2 above threshold all offsets)   b) TFM1 (D^2 above threshold offset 134)
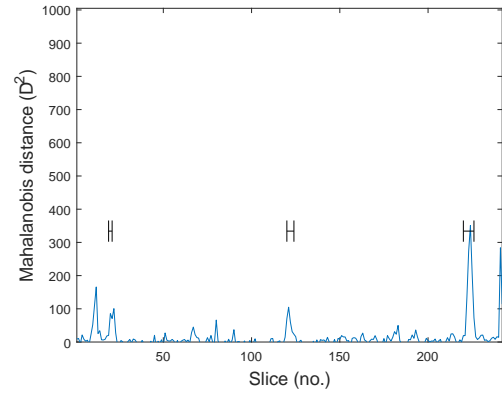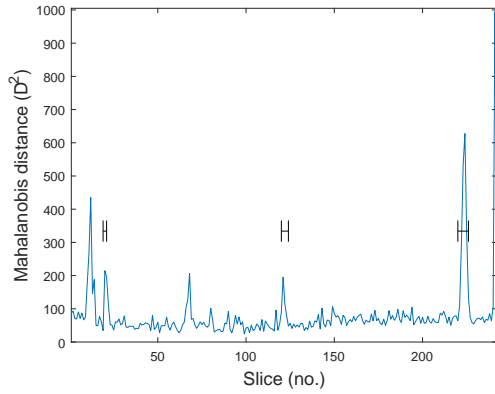
c) TFM2 (max D^2 above threshold all offsets)   d) TFM2 (D^2 above threshold offset 133)
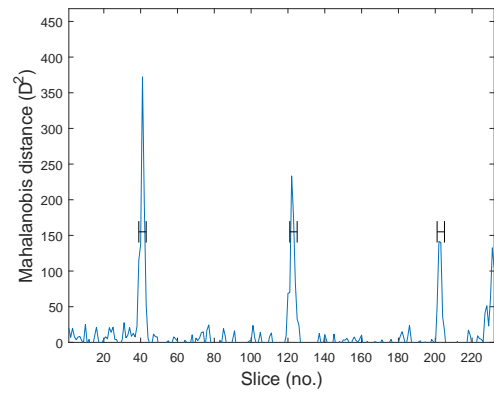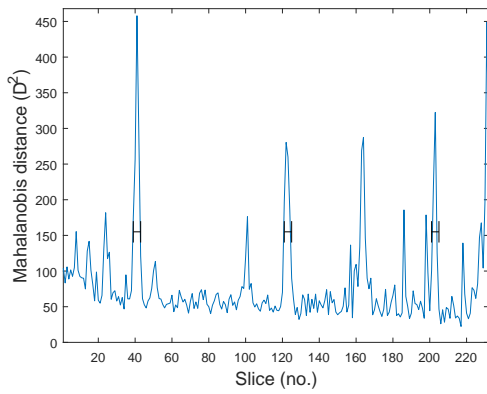
e) TFM3 (max D^2 above threshold all offsets)   f) TFM3 (D^2 above threshold offset 63)
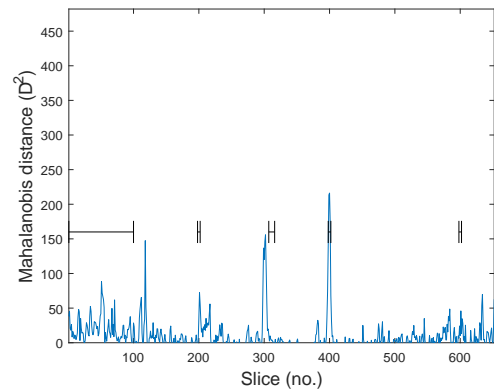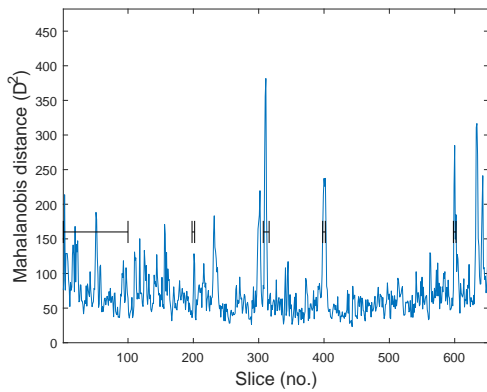
Figure 8.17: Constant offset projections with slice only look-up for each TFM test set

235

there appears to be a lack of discrimination for regions containing a large anomaly.

Test Piece	Kaiser PCs			Average $\chi^2_{(pc,0.95)}$
	Average	Maximum	Minimum	
TFM1	13	17	10	22.4
TFM2	13	16	10	22.4
TFM3	10	13	7	18.31

Table 8.15: Kaiser PCs for each constant offset low rank sub-space using default λ

8.10.4 Constant offset with full image look-up

Images from the constant offset orientation (figure 8.16) contain evidence of most anomalies. Using the full set of projections, where each offset has a threshold set by the respective $\chi^2_{(0.95,pc)}$ confidence level, enables the creation of a binary image for each test piece. A comparison with the respective ground truth template provides the results presented in table 8.17.

The low values of specificity, for the χ^2 confidence level, indicate a high number of false positives for all test pieces, an immediate conclusion being that the confidence level provides too low a threshold. Results from the ROC analysis indicate some potential improvement but sensitivity becomes compromised. Illustrations of these results are provided by figure 8.18.

To aid comparisons the results from the trimming exercise (table 7.28) are duplicated in table 8.18. Overall the results using the default low rank sub-space remain inferior to those from the trimming exercise. In the case of the χ^2 results there is a slight increase in the sensitivity values but the specificities are much reduced. More importantly the specificity values from the ROC analysis never exceed those of the raw χ^2 values obtained after trimming. With one exception (TFM3) the same applies to the sensitivity figures. This and a comparison of the corresponding AUCs leads to the conclusion that under the present conditions PCP is inferior to trimming.

Test Piece	Default λ	PCP SNRs (dB)					Trimming SNRs (dB) (table 7.25)				
		F1	F2	F3	F4	F5	F1	F2	F3	F4	F5
TFM1	0.0643	18.8	15.8	23.0	-	-	42.1	26.1	43.2	-	-
TFM2	0.0657	24.1	20.5	19.0	-	-	45.2	45.1	48.1	-	-
TFM3	0.0392	13.3	12.0	20.1	19.2	21.1	22.7	19.7	25.0	21.2	34.5

Table 8.16: SNRs using default λs

a) TFM1 D^2 above $\chi^2_{(pc,0.95)}$ confidence level

b) TFM1 above ROC threshold

c) TFM2 D^2 above $\chi^2_{(pc,0.95)}$ confidence level

d) TFM2 above ROC threshold

e) TFM3 D^2 above $\chi^2_{(pc,0.95)}$ confidence level

f) TFM3 above ROC threshold
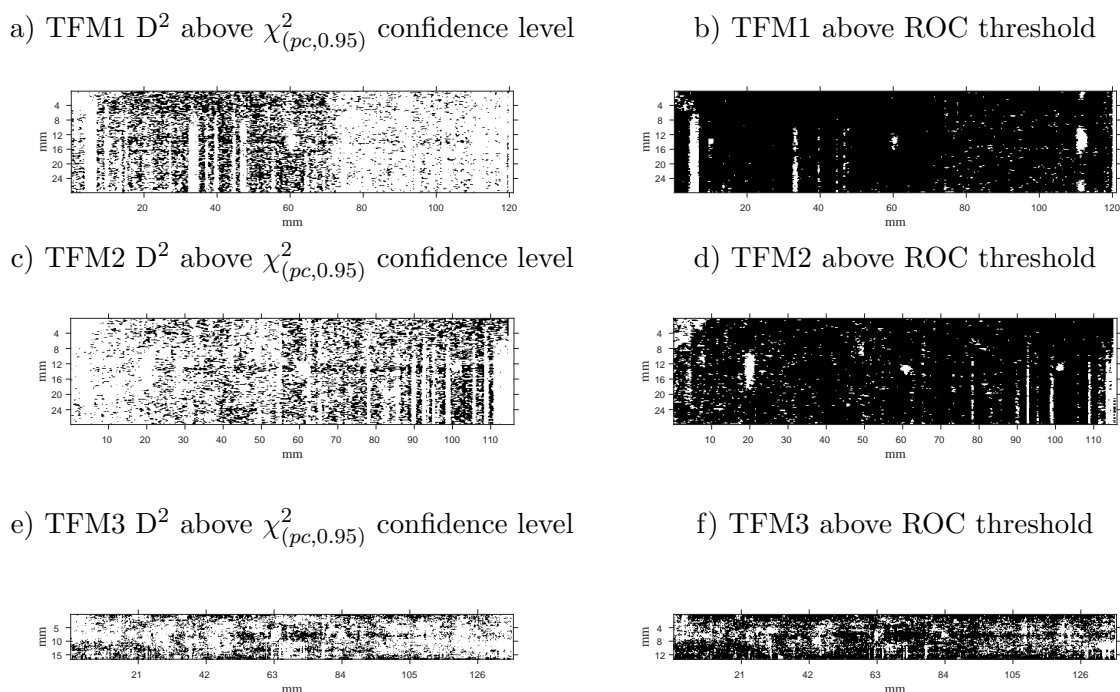
Figure 8.18: TFM images above $\chi^2_{(pc,0.95)}$ and ROC thresholds using $\lambda_{default}$

PCP										
Test Piece	$\chi^2_{(pc,0.95)}$				ROC					
	Acc (%)	Sens.	Spec.	F1	R_m	AUC	Acc (%)	Sens.	Spec.	F1
TFM1	31.2	0.99	0.31	0.05	2.86	0.95	92.7	0.86	0.93	0.07
TFM2	32.3	0.99	0.32	0.01	2.24	0.91	87.7	0.83	0.88	0.02
TFM3	35.4	0.94	0.35	0.03	1.75	0.77	68.3	0.71	0.68	0.05

Table 8.17: ROC values for constant offset with slice and offset look-up

Trimmed training set (from table 7.28)										
Test Piece	$\chi^2_{(pc,0.95)}$				ROC					
	Acc (%)	Sens.	Spec.	F1	R_m	AUC	Acc (%)	Sens.	Spec.	F1
TFM1	89.6	0.99	0.89	0.56	1.36	0.98	96.3	0.98	0.96	0.14
TFM2	90.3	0.91	0.90	0.03	1.13	0.94	95.6	0.90	0.96	0.04
TFM3	86.7	0.86	0.87	0.12	1.01	0.93	85.1	0.88	0.85	0.11

Table 8.18: ROC values for constant offset with slice and offset look-up

Although an increase in the SNR does not necessarily improve overall classification, in this case where all SNRs are comparatively low, an attempt at reducing noise is worthy of

consideration. Section 6.5 demonstrates that de-noising may be achieved by reducing the number of PCs. This can be done by simply reducing the number of Kaiser PCs (table 8.15). In this case, however, an alternative is to reduce the rank of the L matrix. As λ controls the relative weighting between L and S this is done by reducing its value.

8.10.4.1 λ and signal to noise ratio

Figure 8.19 illustrates the variation in SNR of each anomaly for different values of λ. For TFM1 the curves demonstrate a gradual increase as λ is incrementally reduced from its default value. After each peak the SNRs gradually reduce with further reduction in λ. For TFM2 a similar initial behaviour is observed but in this case all peak SNRs correspond to the same λ followed by a sudden reduction. For TFM3 the SNR of all anomalies gradually increases as λ is reduced (from a value of 0.07) to its default value (0.039). With the exception of F1 (the large SDH) all anomalies are at, or close to, their peak SNR.

As previously suggested a higher SNR, although desirable, does not necessarily improve the ability to discriminate between anomaly and background. This is confirmed by tables 8.19, 8.20 and 8.21 which list the classification statistics for each test piece for different λs. The tables also present the result of a ROC analysis to indicate the potential classification.

In terms of the sensitivity and specificity values from the ROC analysis, the results for the default value of λ indicate a better classification than do those for maximum SNRs. Before considering this further some attention is given to TFM3. Here the maximum SNR of most anomalies corresponds to the default value. The exception is the large SDH which has a maximum SNR at $\lambda = 0.02$ (figure 8.19.c). In this case the lower value of λ produces a considerable improvement in specificity with little reduction in sensitivity.

Figure 8.20 now presents the images created from a projection of each test piece using the low rank subspace corresponding to the λs of interest. It is immediately apparent that in each case a reduction in λ produces a reduction in background noise. For TFM1 and TFM2 it is also apparent that the number of anomalous projections with high SNRs are fewer in number leading to a reduction, after thresholding, in sensitivity. For TFM3 the reduced value of λ is chosen to correspond to an increase in the SNR for F1 (the SDH). The improvements in sensitivity and specificity are due largely to the enhancement of this anomaly. As indicated by the image (figure 8.20.f) it is likely that the smaller anomalies will be below any automatically derived threshold.

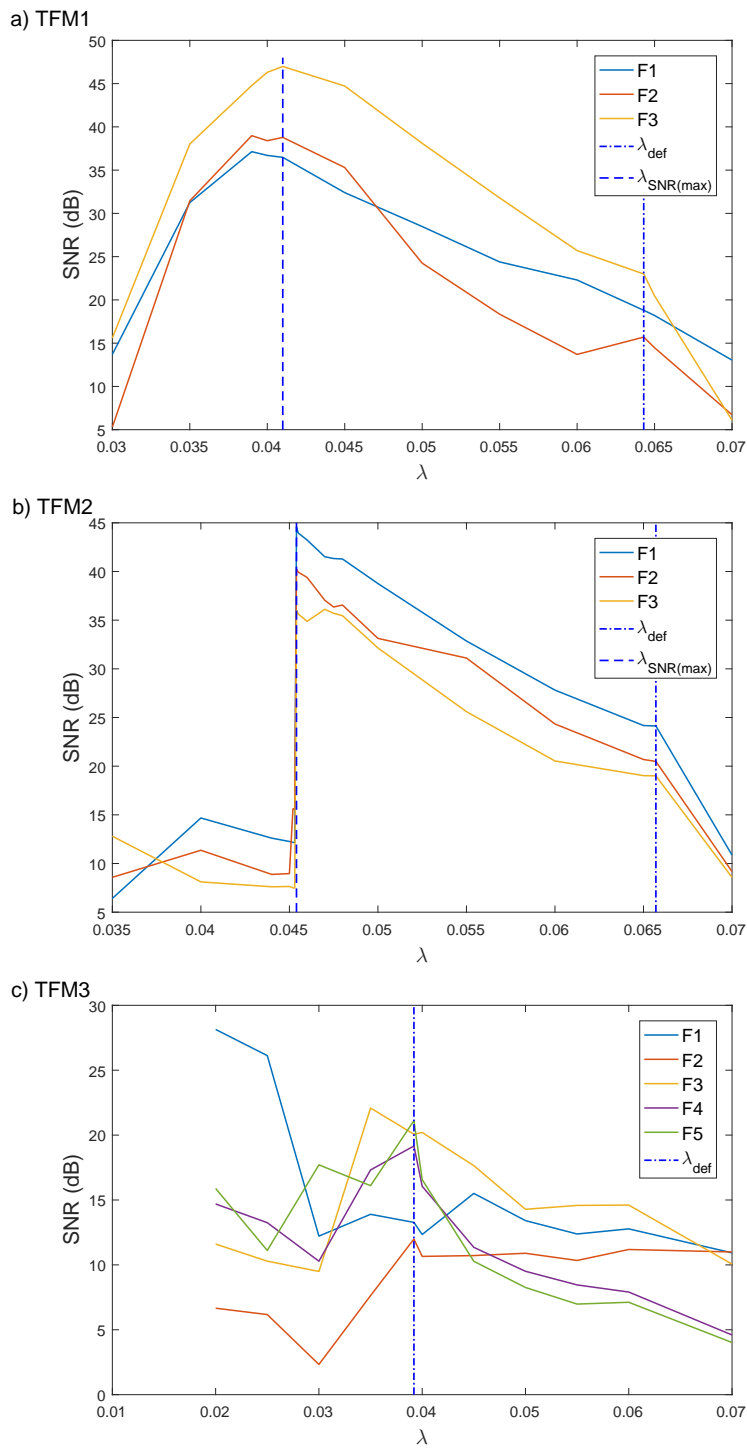These observations indicate that, for the TFM images, although the default value of λ is
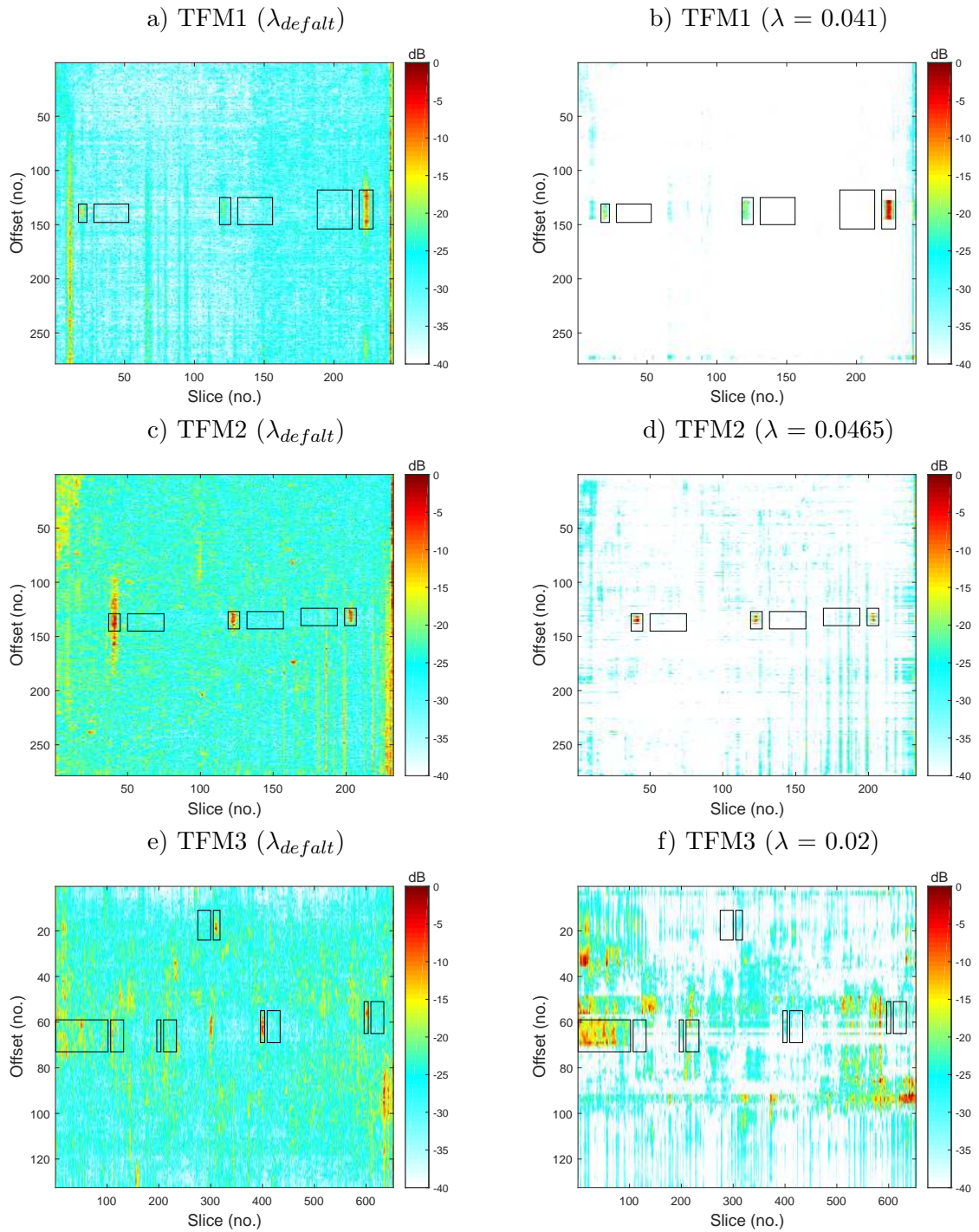
238

Figure 8.19: SNR of each fault versus λ

a) TFM1 (λ_{defalt})

b) TFM1 ($\lambda = 0.041$)

c) TFM2 (λ_{defalt})

d) TFM2 ($\lambda = 0.0465$)

e) TFM3 (λ_{defalt})

f) TFM3 ($\lambda = 0.02$)

Figure 8.20: TFM images using $\lambda_{default}$ and λ for maximum SNR.

	λ	χ^2			ROC					Rank
		Acc (%)	Sens.	spec.	AUC	Acc(%)	Sens.	Spec.	F1	(mean)
	0.03	74.3	0.54	0.74	0.68	65.7	0.60	0.66	0.01	3
	0.035	73.9	0.68	0.74	0.80	83.6	0.67	0.84	0.02	5
	0.04	76.2	0.81	0.76	0.85	93.4	0.77	0.93	0.07	10
SNR_{max}	0.041	77.6	0.82	0.78	0.87	92.4	0.79	0.92	0.06	11
	0.045	76.5	0.98	0.76	0.95	94.9	0.95	0.95	0.10	18
	0.05	62.1	0.99	0.62	0.98	94.6	0.93	0.95	0.09	30
	0.055	47.5	0.99	0.47	0.97	92.8	0.93	0.93	0.07	42
	0.06	35.1	0.99	0.35	0.96	90.2	0.90	0.90	0.05	54
default	0.0643	31.3	0.99	0.31	0.95	93.7	0.85	0.94	0.08	64
	0.07	92.3	0.85	0.92	0.95	90.3	0.87	0.90	0.05	77
	0.08	43.0	0.98	0.43	0.93	87.1	0.87	0.87	0.04	96
	0.65	92.3	0.85	0.92	0.92	89.8	0.83	0.90	0.05	112

Table 8.19: TFM1 - ROC analysis results

	λ	χ^2			ROC					Rank
		Acc (%)	Sens.	Spec.	AUC	Acc(%)	Sens.	Spec.	F1	(mean)
	0.035	79.3	0.15	0.79	0.51	41.9	0.71	0.42	0.004	5
	0.04	79.6	0.15	0.79	0.49	46.9	0.58	0.47	0.003	9
	0.045	75.9	0.17	0.76	0.47	36.5	0.63	0.37	0.003	17
	0.046	73.6	0.39	0.74	0.60	56.4	0.54	0.56	0.004	20
SNR_{max}	0.0465	72.0	0.53	0.72	0.67	67.0	0.55	0.67	0.005	21
	0.0475	1.0	1.0	0.0	0.80	71.4	0.76	0.71	0.008	21
	0.05	63.7	0.76	0.63	0.81	85.8	0.70	0.86	0.014	27
	0.055	48.6	0.90	0.48	0.87	86.3	0.80	0.86	0.017	38
	0.06	37.2	0.94	0.37	0.87	91.6	0.73	0.92	0.024	50
	0.065	32.4	0.98	0.32	0.91	83.9	0.84	0.87	0.018	62
default	0.0657	32.5	0.98	0.32	0.91	87.7	0.83	0.88	0.019	63
	0.07	92.6	0.83	0.93	0.89	89.2	0.82	0.89	0.021	70
	0.08	42.3	0.96	0.42	0.91	93.0	0.84	0.93	0.034	91
	0.09	58.9	0.94	0.58	0.92	94.1	0.84	0.94	0.039	104
	0.1	73.6	0.94	0.74	0.93	92.7	0.85	0.93	0.032	109
	0.2	92.0	0.83	0.92	0.91	91.0	0.83	0.91	0.025	109
	0.3	92.5	0.83	0.93	0.90	90.2	0.82	0.90	0.024	109
	0.4	92.6	0.83	0.92	0.90	90.4	0.82	0.90	0.024	109
	0.65	92.6	0.83	0.93	0.89	89.2	0.82	0.89	0.021	109

Table 8.20: TFM2 - ROC analysis results

not optimum, in terms of maximising sensitivity and specificity, it does represent a value that is in keeping with the maximums achieved. It is also indicated that whilst an anomaly must be distinct from its background, maximising its SNR does not necessarily lead to an improved classification.

		χ^2			ROC					Rank
	λ	Acc (%)	Sens.	Spec.	AUC	Acc(%)	Sens.	Spec.	F1	(mean)
	0.02	59.7	0.94	0.59	0.91	85.5	0.83	0.86	0.11	3
	0.025	60.4	0.90	0.61	0.83	77.5	0.76	0.78	0.07	7
	0.03	49.8	0.87	0.49	0.78	73.6	0.70	0.74	0.05	18
	0.035	39.4	0.89	0.39	0.71	62.5	0.68	0.62	0.04	44
default	0.0392	35.4	0.94	0.34	0.77	68.3	0.71	0.68	0.05	45
	0.04	35.2	0.95	0.34	0.79	69.5	0.74	0.70	0.05	47
	0.045	42.2	92.9	0.42	0.80	71.5	0.73	0.72	0.05	62
	0.05	54.8	0.89	0.54	0.82	85.9	0.72	0.76	0.06	74
	0.055	68.5	0.83	0.68	0.83	76.7	0.75	0.77	0.07	74
	0.06	76.7	0.77	0.77	0.83	76.0	0.76	0.76	0.06	74
	0.065	81.3	0.74	0.81	0.84	75.1	0.77	0.75	0.06	74
	0.4	91.2	0.49	0.92	0.82	78.0	0.71	0.78	0.07	76

Table 8.21: TFM3 - ROC analysis results

For TFM3 the situation is not as straightforward. This is largely due to the size differences between the large SDH and the smaller SBHs when viewed, as in the images, from above. This suggests some possibility of tuning λ to detect anomalies of a particular size.

8.10.5 Reduced PCs and classification

To complete this investigation results are now presented for a reduction in the number of PCs. As maximising the SNR is shown not to improve the classification these comparisons continue without recording the resulting changes to SNRs. The results are, instead, recorded directly as changes in classification statistics.

After PCP the number of Kaiser PCs is already small (table 8.15). This comparison is, therefore, limited to only two reductions (KaiserPCs/2 and KaiserPCs/1.5). Classification results using these, with the default value of λ, are listed in table 8.22. With the exception of a slight improvement in sensitivity for TFM1 there are no improvements of significance.

The conclusion is that variations to either the default λ or number of PCs do not produce significant improvements to the classification statistics. Consequently and with no alternative methods of calculation, for this application, the recommended value of λ is that determined by equation 8.4 and the number of PCs is that determined by the Kaiser stopping rule.

ROC analysis (tables 8.19, 8.20 and 8.21) shows that in all cases a threshold above the χ^2 confidence level can produce significant improvements. However a comparison with

Test piece	χ^2			ROC					PCs Used
	Acc (%)	Sens.	Spec.	AUC	Acc(%)	Sens.	Spec.	F1	
TFM1	31.3	0.99	0.31	0.95	93.7	0.85	0.94	0.08	Kaiser
	46.3	0.99	0.46	0.96	89.8	0.92	0.90	0.05	Kaiser/1.5
	63.9	0.98	0.64	0.96	92.9	0.91	0.93	0.07	Kaiser/2
TFM2	32.5	0.98	0.32	0.91	87.7	0.83	0.88	0.02	Kaiser
	47.2	0.94	0.47	0.88	88.4	0.76	0.88	0.02	Kaiser/1.5
	64.1	0.88	0.64	0.86	82.3	0.80	0.82	0.01	Kaiser/2
TFM3	35.4	0.94	0.34	0.77	68.3	0.71	0.68	0.05	Kaiser
	49.9	0.88	0.49	0.78	71.0	0.69	0.71	0.05	Kaiser/1.5
	64.5	0.77	0.64	0.75	68.7	0.69	0.69	0.05	Kaiser/2

Table 8.22: TFM classifications for different PCs

trimming (8.18) indicates that PCP remains inferior.

8.10.5.1 Additional thresholding

With the previous acknowledgment, and in the interest of completion, the results of additional thresholding, using the KI, Otsu and ME techniques, are listed in table 8.23.

In all instances the threshold is greater than the ROC recommendation. Consequently the sensitivity is low and, more significantly, the overall classification does not approach that for trimming. Accepting that no technique produces satisfactory results it is noted that on occasions the Otsu threshold produced higher sensitivity figures than the KI threshold. The reason for this is that now the background level (figure 8.20) is more significant than in previous cases. One consequence of this is that although the histogram remains largely unimodal, on some occasions a bimodal pattern begins to emerge.

8.10.6 Blob detection

For the available TFM images anomaly detection by PCP is not as attractive as by trimming. In addition to comparatively poor figures for sensitivity and specificity the AUCs indicate a generally better classification using trimming. For consistency with the results for trimming, visual demonstrations of the effect of various thresholds are presented. Although such images may be of assistance to an inspector of more importance is the ability to extract size and location information relating to potential anomalies. Presently this is achieved by the process of blob detection.

Test Piece	λ	Method	R_m	Acc.(%)	Sens.	Spec.	F1
TFM1	0.0643	D^2	1	31.3	0.99	0.31	0
		$D^2_{(ROC)}$	2.68	92.7	0.86	0.93	0.07
		KI	3.73	97.0	0.72	0.97	0.13
		Otsu	6.53	98.8	0.52	0.99	0.22
		ME	7.89	99.1	0.42	0.99	0.23
TFM2	0.0657	D^2	1	32.5	0.98	0.32	0
		$D^2_{(ROC)}$	2.19	87.7	0.83	0.88	0.02
		KI	3.56	97.2	0.63	0.97	0.05
		Otsu	3.00	95.4	0.75	0.95	0.05
		ME	5.11	98.8	0.46	0.99	0.1
TFM3	0.0392	D^2	1	35.4	0.94	0.34	0
		$D^2_{(ROC)}$	1.69	68.3	0.71	0.68	0.05
		KI	4.40	97.4	0.15	0.98	0.11
		Otsu	2.11	80.4	0.57	0.81	0.06
		ME	8.35	98.8	0.02	0.99	0.04
TFM3	0.02	D^2	1	59.7	0.94	0.59	0
		$D^2_{(ROC)}$	0.66	85.5	0.83	0.86	0.11
		KI	1.17	93.4	0.67	0.94	0.18
		Otsu	1.46	95.1	0.60	0.96	0.21
		ME	3.64	98.4	0.14	0.99	0.16

Table 8.23: Classification statistics for different thresholds

All figures (8.24 to 8.27) are presented at the end of this chapter. When using the default λ all anomalies, with the exception of TFM3's SDH, are evident after applying the KI threshold. For this particular instance the Otsu threshold is more effective although it also produces considerably more false positives. For $\lambda = 0.02$ the KI and Otsu thresholds both detect TFM3's SDH but the smaller SBHs are now less distinct. This contrasts strongly with trimming where all anomalies are detected by the KI threshold.

The results of blob detection after post processing (section 7.10.3.1) are illustrated in figure 8.21. Once again there is a contrast with the same exercise after trimming (figure 7.22). Here the images demonstrate the difficulties of detecting anomalies of different sizes. Although the default value of λ is satisfactory for TFM1 and TFM2, where all anomalies are of a similar size, detection of TFM3's SDH is less emphatic. It is also noted that F2 (a ϕ1mm SBH) is missed. With λ set to 0.02 (figure 8.21.d) there is a more emphatic indication of the SDH but now many of smaller SBHs are missed.
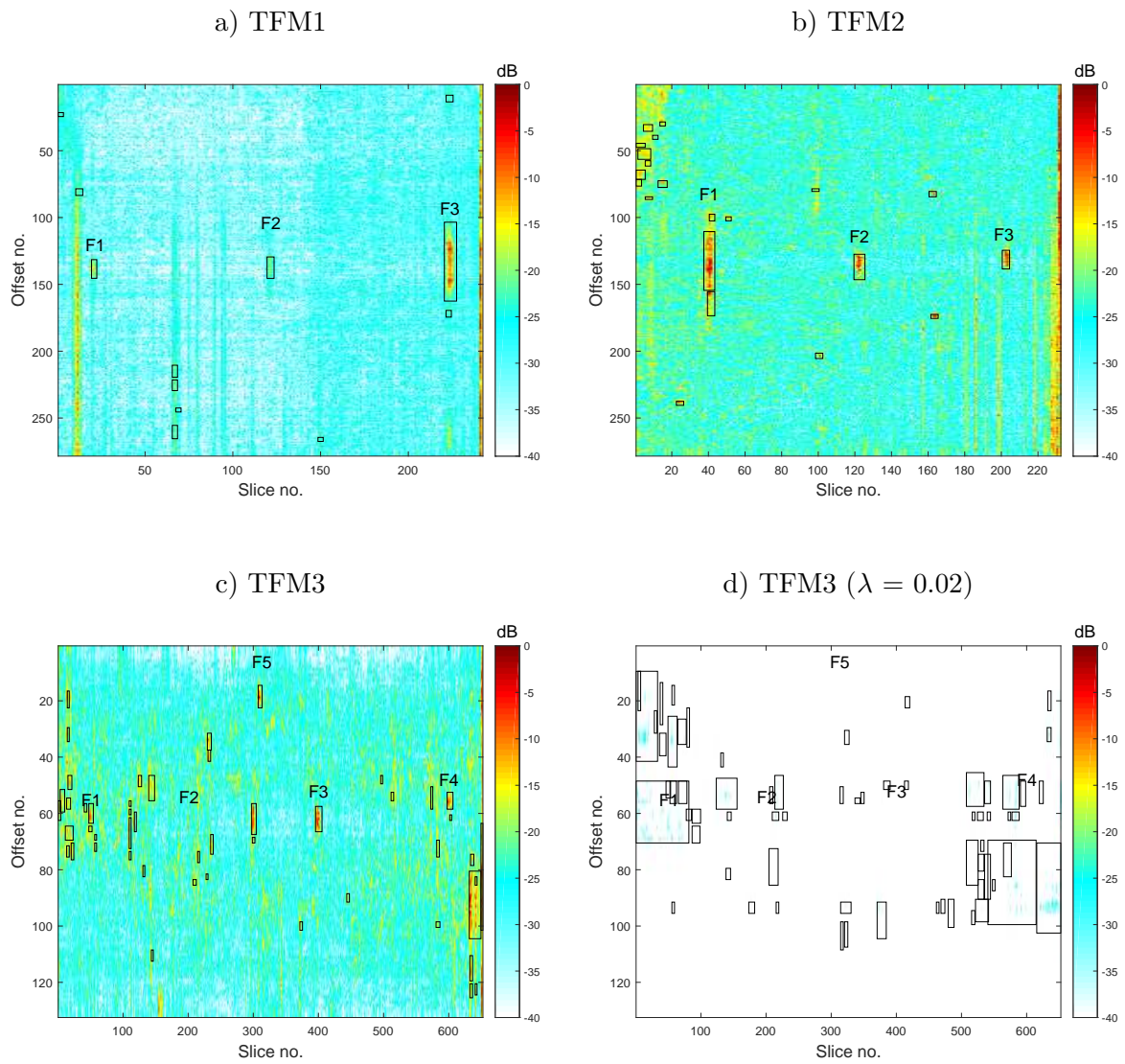
Figure 8.21: TFM blob detection (default λ)

8.10.6.1 Sparse matrix

So far the emphasis has been to find an accurate low rank representation of the background. Once found, this can be repeatedly used to efficiently detect outliers in any number of instances of the same weld type. In this context the sparse matrix is largely considered as a by product of the PCP algorithm. However as a repository of non-background information the sparse matrix contains a description of outliers in the reference set. It is emphasised that, unlike the low rank matrix which is determined once and used multiple times, a new sparse matrix needs to be created for each weld. Consequently computational expense would likely preclude its use in practice. However, to complete this investigation some attention is now given to the sparse representation.

Unlike the low rank matrix, which ideally represents an accurate description of the undisturbed background, the sparse matrix contains a description of any disturbances. A visual comparison of the two is given by figure 8.2. The example is for a single offset selected from TFM2 because it is known to pass through all three FBHs. The more prominent features of the low rank matrix, figure 8.2.b, are the front and back walls. In the sparse matrix, figure 8.2.c, these background features are diminished with the three FBHs becoming more evident.

For this cursory evaluation the full set of S matrices (one for each offset) are condensed into a single matrix representative of a plan view of the test piece. This is achieved by taking the maximum value from each column (representing depth through the test piece) of the original 3D structure. To distinguish this matrix from the individual sparse matrices it is denoted as S_{max}. Similarly a low rank matrix (L_{max}) is created. Using the default value of λ, figure 8.22 provides an illustration of S_{max} and L_{max}, for each test test piece.

As expected the low rank image contains very little variation; it is largely representative of the background and shows little evidence of anomalies. By contrast an immediate observation of the sparse representations is that many of the anomalies are now clearly evident. These are, however, accompanied by other artifacts as well as general background noise. With the same areas used earlier, table 8.24 lists the SNRs for each fault. For the default λ these values are inferior to those acquired using projections against the low rank subspace, table 8.16.

For demonstration the same SNR calculations are repeated for the low rank representations, table 8.25. As expected these are much lower than those for the sparse representation indicating no significant disturbance in the background.
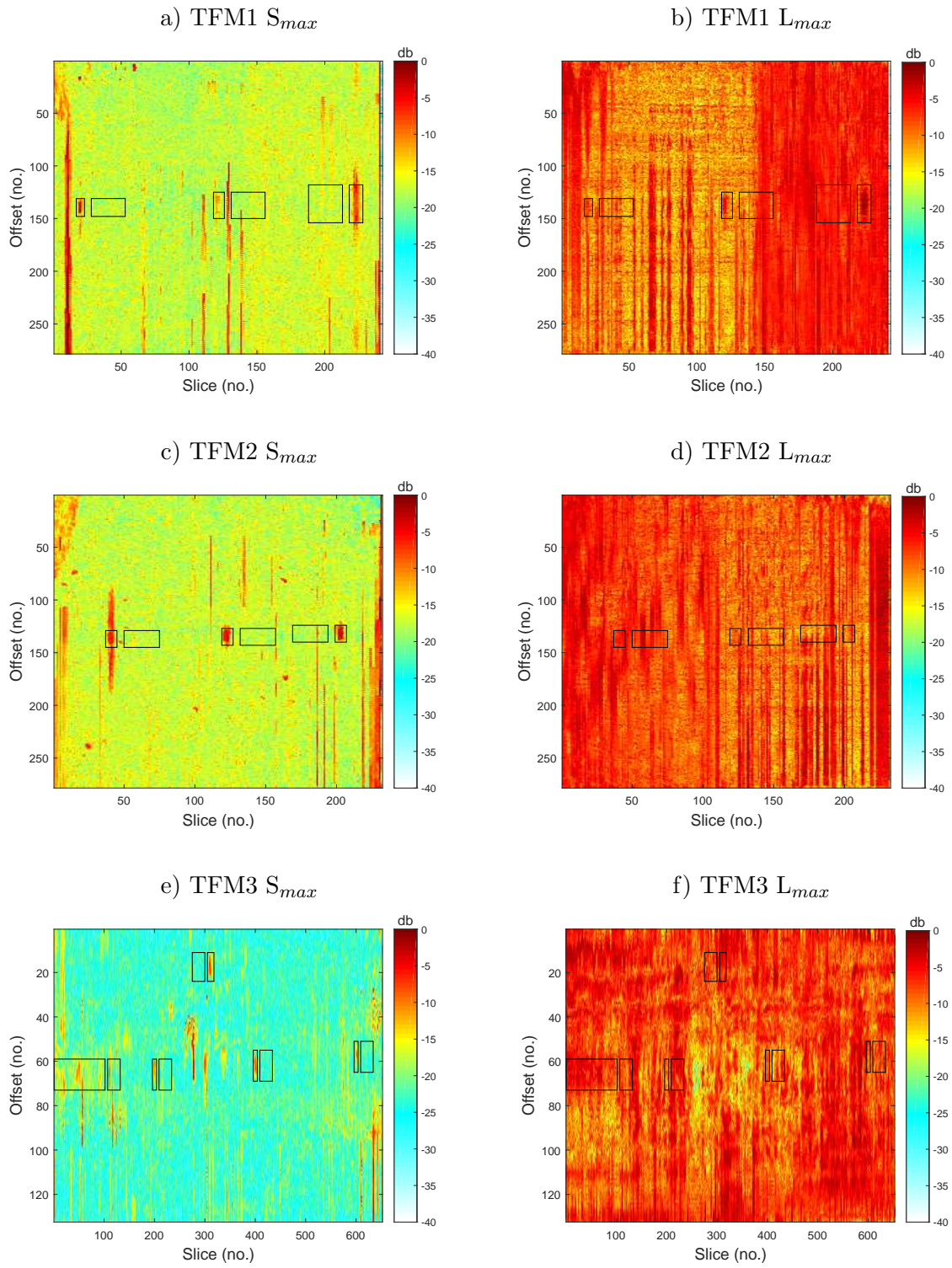
a) TFM1 S$_{max}$

b) TFM1 L$_{max}$

c) TFM2 S$_{max}$

d) TFM2 L$_{max}$

e) TFM3 S$_{max}$

f) TFM3 L$_{max}$

Figure 8.22: TFM images for default λs

247

Test piece	SNRs (dB)				
	F1	F2	F3	F4	F5
TFM1	14.7	7.1	13.1	-	-
TFM2	11.7	14.4	14.7	-	-
TFM3	10.2	12.8	14.7	13.7	16.1

Table 8.24: Sparse (S_{max}) SNRs

Test piece	SNRs (dB)				
	F1	F2	F3	F4	F5
TFM1	4.3	3.8	6.1	-	-
TFM2	4.1	6.2	5.1	-	-
TFM3	6.8	0.2	3.4	4.7	5.3

Table 8.25: Low rank (L_{max}) SNRs

Test Piece	λ	Method	AUC	Acc.(%)	Sens.	Spec.	F1
TFM1	0.0643	$D^2_{(ROC)}$	0.91	88.2	0.81	0.88	0.06
		KI		97.1	0.62	0.97	0.12
		Otsu		97.6	0.54	0.98	0.13
		ME		96.9	0.66	0.97	0.12
TFM2	0.0657	$D^2_{(ROC)}$	0.92	92.1	0.87	0.92	0.02
		KI		98.2	0.61	0.98	0.09
		Otsu		92.9	0.84	0.93	0.03
		ME		96.5	0.70	0.97	0.05
TFM3	0.0392	$D^2_{(ROC)}$	0.84	74.9	0.77	0.75	0.06
		KI		97.2	0.17	0.98	0.12
		Otsu		95.8	0.27	0.97	0.12
		ME		98.3	0.03	0.99	0.04
TFM3	0.02	$D^2_{(ROC)}$	0.85	76.2	0.80	0.76	0.07
		KI		97.5	0.15	0.98	0.11
		Otsu		95.1	0.31	0.96	0.12
		ME		98.2	0.06	0.99	0.07

Table 8.26: ROC and threshold classification statistics from S_{max}

Despite the desirability of a larger SNR this does not by itself necessarily improve the probability of detection. To complete this comparison consideration is given to the results of ROC analysis, thresholding (KI, Otsu and ME) and finally blob detection.

Table 8.26 presents the results of the ROC analysis and image thresholding. In comparison

a) TFM1 ($\lambda_{def.}$)

b) TFM2 ($\lambda_{def.}$)

c) TFM3 ($\lambda_{def.}$)

d) TFM3 ($\lambda = 0.02$)
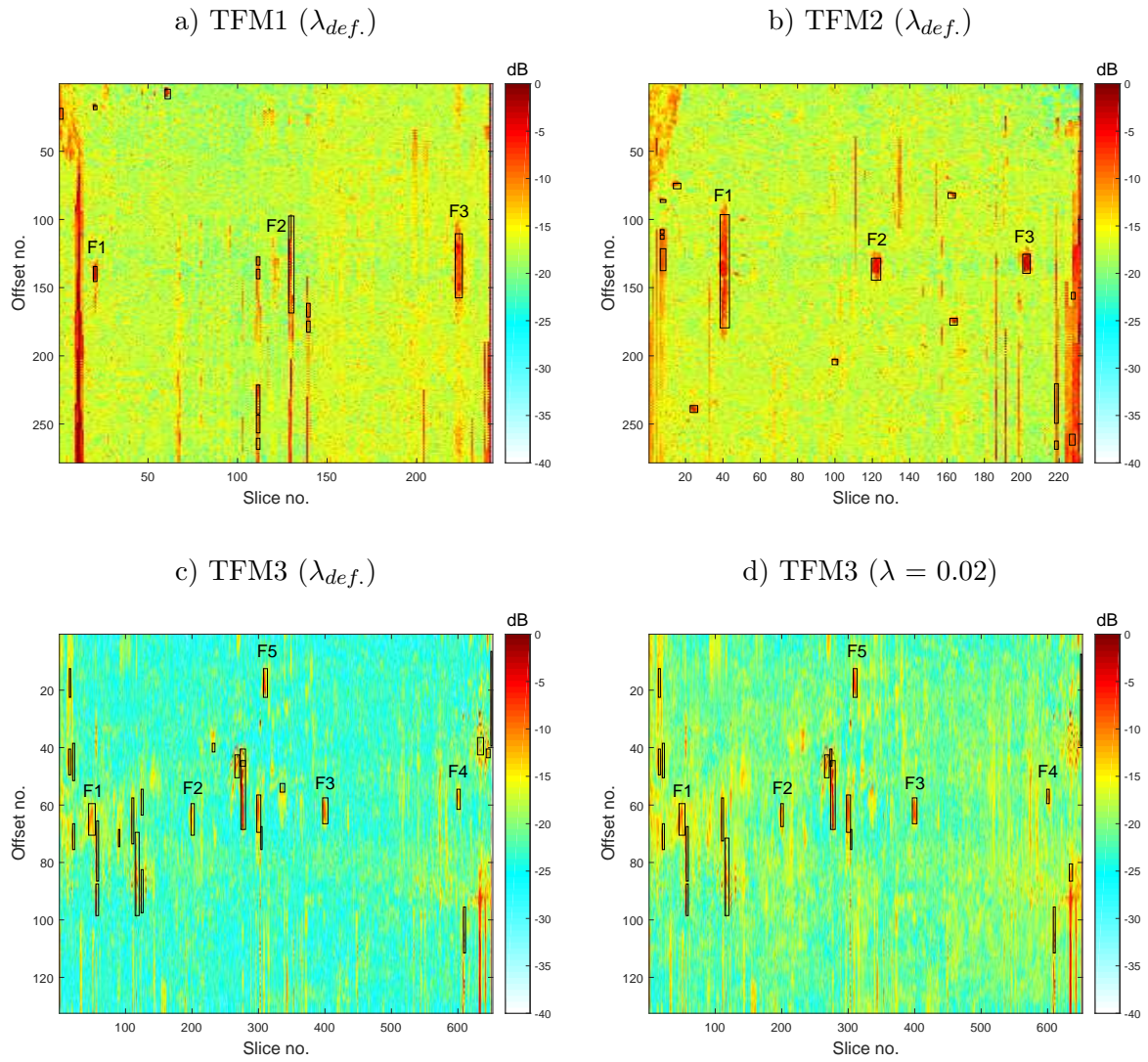
Figure 8.23: Blob detection of S_{max} after post processing

249

with the same test using the low rank subspace (table 8.23) results for TFM2 and TFM3 (with default λ) are an improvement; the opposite is the case for TFM1 and TFM3 with reduced λ.

As a final comparison the results of blob detection, after post processing, are illustrated in figure 8.23. Once again these have many similarities with results of the same exercise with projections using the low rank sub-space, figure 8.21. The most significant difference is that in this case F2, in TFM2, is not detected.

This exercise has demonstrated some potential application for the sparse matrix. Despite the differences in classification these are slight with no significant overall improvement. On the contrary, blob detection fails to detect one anomaly. Whilst the low rank sub-space may be used repeatedly for detection of anomalous conditions of many instances of the same type of weld, a new sparse matrix must be created for each new instance. This is a time consuming operation.

8.10.7 Chapter summary

Robust PCA using PCP has been applied to both the sectorial and TFM images. In both cases the weighting between the L and S matrix, λ, is found to have a significant influence on the results.

For the sectorial images a default value (λ_{def}) from equation 8.4 proved unsatisfactory. Through an iterative process a value of $10\times\lambda_{def}$, was found to produce a considerable improvement for both the full sector and constant A-scan orientations. Although the reason for this is not known, both orientations produce high dimensional data sets. By comparison the dimensionality of the TFM constant offset orientation is much lower and here default λ proved to be satisfactory.

After doubling the number of PCs, results using the full sector orientation produced results similar to those for trimming alone. The greatest improvement occurs with the constant A-scan orientation using sector and A-scan look-up. Apart from increasing the default λ no further tuning was required. The full potential of the approach is demonstrated by the ROC analysis (table 8.11). Although these results represent a theoretical potential this is almost achieved using KI thresholding (table 8.13). The resulting false positives are likely to be areas requiring further exploration. If these results are be reproduced over a wider range of data sets the approach has considerable potential.

For the TFM images the full image orientation did not produce reliable results and was

quickly discounted as a method of anomaly detection. The effectiveness of the constant offset orientation with slice and offset look-up is not clear. ROC analysis indicates that some classification is possible but the potential is inferior to that for trimming. There are a number of possible reasons for this, all leading to suggestions for further work. The following comments apply only to the constant offset orientation with slice and offset look-up.

Initially the intention was to produce a set of low rank matrices directly from each TFM offset with each test observation only to be mean centred before projection. Resulting images contained indications of the induced anomalies but ROC analysis indicated very poor potential for classification. Improved results were obtained using standardised data. Despite this the ROC analysis indicates PCP to be inferior to trimming (tables 8.17 and 8.18).

Changes to λ and the number of PCs to use were the main areas of investigation. Although a reduction in λ increased the SNR of individual anomalies, this did not necessarily improve the probability of detection as determined by sensitivity. The eventual conclusion was that the default λ and number of PCs (determined by the Kaiser stopping rule) were satisfactory, an additional reason for this being that no alternative method of estimation, other than trial and error, is available.

All projection distances using standardised data remain in a relatively small range. This applies to many, but not all, projections using mean centred data. As in the case of trimming there are also a number of projections with unusually high projection distances. The difference now is that, in the case of the mean centred projections the number, and magnitude, of these excessive projections is too large for effective thresholding. One possible reason for this is that TFM images are created through a number of mathematically complex operations on the underlying FMC data. Although the operations on each slice are the same, slight differences potentially lead to larger changes in the eventual image. This is most evident where there is a tilting between two or more adjacent images. It is also noted that many excessive projections correspond to these disturbances.

Sectorial data sets consist of A-scans that are not subject to further mathematical processing. The superiority of these results compared with those for TFM images strongly suggests that, for further investigation, PCP be applied directly to the FMC data set.

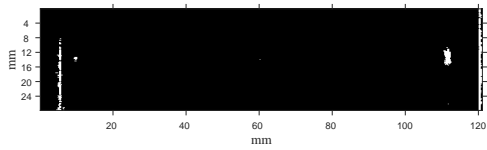a) D^2 above $\chi^2_{(pc,0.95)}$ confidence level
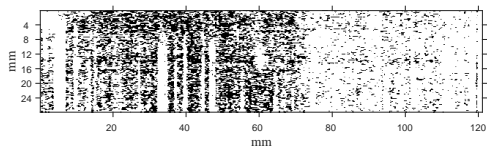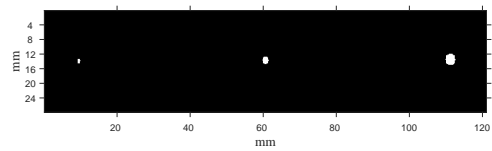
b) Ground truth
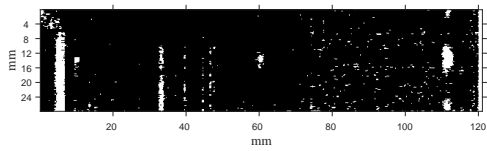
c) ROC threshold
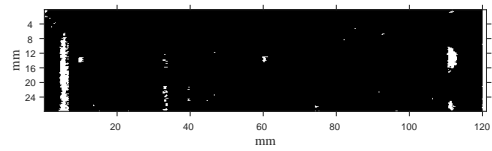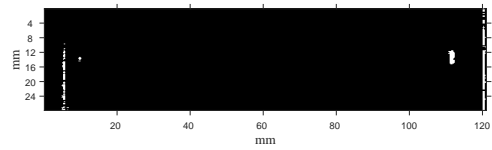
d) KI threshold

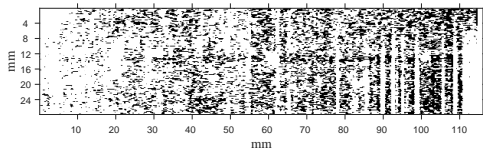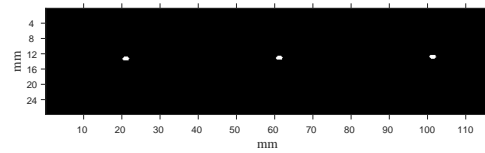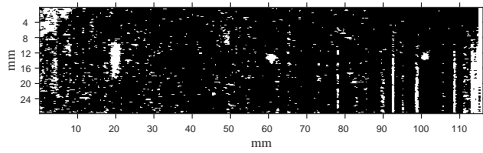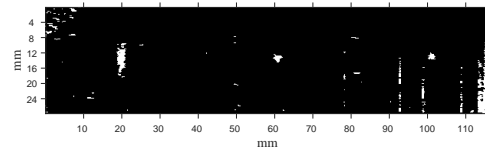e) Otsu threshold

f) ME threshold

Figure 8.24: TFM1 D^2 projections with thresholding $\lambda_{default}$

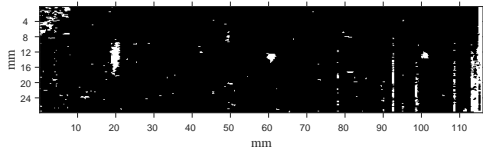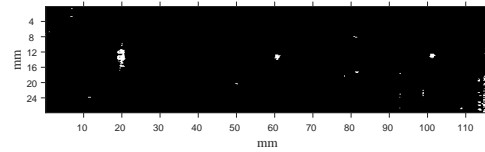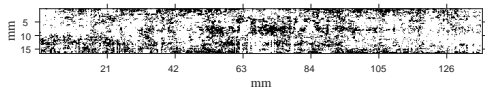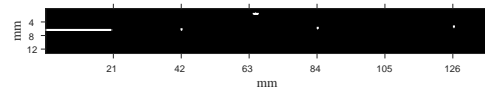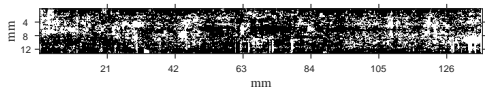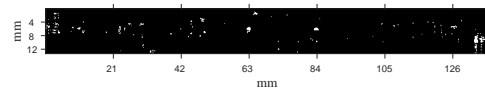a) D^2 above $\chi^2_{(pc,0.95)}$ confidence level

b) Ground truth

c) ROC threshold

d) KI threshold

e) Otsu threshold

f) ME threshold

Figure 8.25: TFM2 D^2 projections with thresholding $\lambda_{default}$

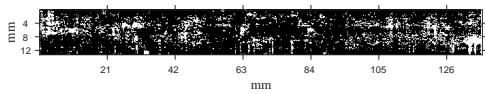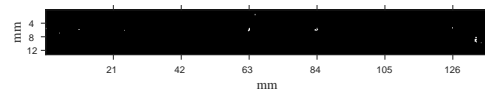a) D^2 above $\chi^2_{(pc,0.95)}$ confidence level

b) Ground truth

c) ROC threshold

d) KI threshold

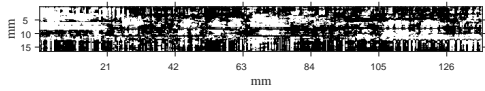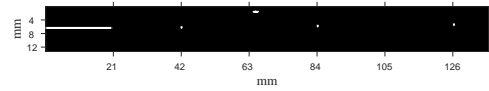e) Otsu threshold

f) ME threshold

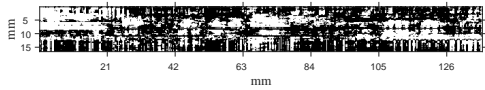Figure 8.26: TFM3 D^2 projections with thresholding $\lambda_{default}$

a) D^2 above $\chi^2_{(pc, 0.95)}$ confidence level
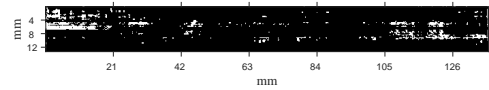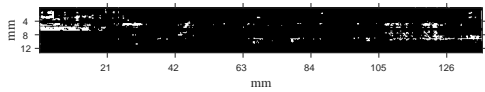

b) Ground truth

c) ROC threshold

d) KI threshold

e) Otsu threshold
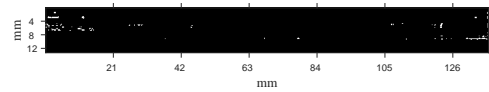

f) ME threshold

Figure 8.27: TFM3 D^2 projections with thresholding $\lambda = 0.02$

Chapter 9

Conclusions

Today AUT driven inspection processes tend to use traditional, single element probes, to inspect selected zones of the weld. Replacing these with phased arrays gives the AUT system an ability to inspect the weld's entire volume thereby reducing the probability of missing a defect. This does, however, present the inspector with hundreds of individual cross sectional images for each weld, further adding to time pressure and mental workload. The primary objective of this work is, therefore, to efficiently and reliably identify regions of a weld containing potential anomalies. These are brought to the attention of the inspector for sentencing, with automation of this step being a subject for future work.

To address the objective a first step in this approach is to emulate the method used in radiography. Here a single radiograph presents a plan view along a length of the weld. In essence this compresses the data into a 2 dimensional image. The resulting image has similarities to a conventional C-scan which provides a map from peak A-scan values. It is found that, although this univariate approach identifies larger anomalies, small features of interest have little evidence. Subsequently there is a poor probability of detection. Another limitation of this when using cross sectional image data, is that it is only possible in the absence of a front wall.

In place of a conventional C-scan, PCA is found to provide a more reliable indication of anomalies. Two significant findings of this work are that, when used with PCP, the results from sectorial scans produce a considerable improvement in the probability of detection. In the case of immersion TFM images, an equally significant finding was that the front wall becomes no hindrance to an observation of the underlying structure.

Limitations of PCA due to outliers have been mitigated by the use of data trimming and

the application of PCP. In practice other problems of PCA relate to the number of principal directions (or principal components) needed for satisfactory classification. To this end an objective has always been to seek a simple method of tuning. For example, a particular type of information may be contained within a range of principal directions not necessarily starting from the first PC. Determining an optimum range often requires a number of trial and error attempts and is not suitable for automation. Here, if a technique does not produce a promising result, using the principal components, and starting at 1, it has been rejected.

Similarly, deciding how many directions to use is subjective. Once again, the pragmatic decision has been to favour a method that provides an acceptable result over an alternative more optimum solution, provided it may be automated. In many of these applications the Kaiser stopping rule was found to provide an acceptable solution.

The two dimensional image, resulting from peak D^2 distances, is subject to thresholding and blob detection. This makes possible the 2D sizing and location of anomalies. Subsequent comparisons with ground truth information provides quantitative results relating to the probability of detection (sensitivity) and resilience to false positives (specificity). Although this classification provides no depth information, the full volume relating to each anomaly can be retrieved using the 2D location information as a key into the original 3D data set. This now provides the data for further classification and sentencing.

The results show that with a representative subspace high accuracy results are achievable. This is most pronounced in the case of the PCP derived subspace of the sectorial data. In particular table 8.11 indicates that without additional thresholding a POD of 100% with a false alarm rate of only 5% is possible. However PCP has proved to be computationally time consuming and, for these applications, a method of deriving parameters (such as the weighting factor, error limit and number of PCs to use) is not clear. It has not, for example been possible to produce similar results for the TFM data sets.

An alternative to PCP, data set trimming, is far simpler. Although results for sectorial data are inferior to those using PCP they are, nonetheless, encouraging. For example table 7.16 illustrates that after KI thresholding the trimmed subspace of the sectorial data produces an average POD greater than 94% with an average false alarm rate less than 9%. The table also indicates that these are close to the 'optimum' determined by ROC analysis. For the TFM data sets table 7.34 indicates that, without further thresholding, data set trimming produces an average POD of 92% with an average FAR of 10%.

In terms of anomaly detection the POD rates from trimming are actually better than the

above values suggest. For example despite the POD not reaching 100% all anomalies are identified by at least 50% of the respective pixels registering as true positives. Classification statistics for the trimming example are, therefore, conservative. Actual performance, in terms of identifying a region containing an anomaly, is better than suggested by the PODs quoted here (and derived from tables 7.16 and 7.34). Further improvements, particularly in the False Alarm Rate (FAR) values can be made by simple post processing as illustrated in figure 7.21, and described in section 7.10.3.1.

Further work, in terms of establishing a representative anomaly free subspace, is required. With limited data sets this work has demonstrated that PCA provides considerably improved results over those using peak A-scan or, in the case of TFM images, peak column values. It also demonstrates that for immersion tests the presence of a dominating front wall does not hinder automatic anomaly detection.

As cross sectional data from ultrasonic AUT systems becomes progressively more common, so too does the importance of automating the inspection process. The work presented here demonstrates that such automation is viable for welds of a consistent geometry. Initially, confidence in such a system can be gained by comparing results of manual inspection to those achieved using the methods developed here. With this confidence, adoption of such a system has the potential to vastly reduce inspection time whilst producing a high POD. Although false alarm rates will be present, they will be low and will likely represent borderline conditions requiring the knowledge and expertise of an experienced inspector for final sentencing.

9.1 Contribution to knowledge

This work is believed to represent the first attempt at using PCA as a means of identifying anomalies in ultrasonic data. This statement is qualified by acknowledging that previous researchers have used PCA to characterise faults that exist in single test pieces and at known locations. Here the objective is to locate anomalies in the potentially tens of thousands of images produced on an industrial scale by modern techniques of AUT. Characterisation of these anomalies becomes the objective of further analysis not covered here.

The novel contributions of this work to the detection of anomalous regions in ultrasonic images are:-

- The application of PCA for anomaly detection in an indexed sequence of ultrasonic

images. Specifically for anomaly detection in large sequences of such images, this is thought to be the first investigation using multivariate techniques.

- The development of a multivariate method for suppressing the dominance of the front (and back wall) background features, particularly in the case of an immersion test. This enables the creation of a plan view image (analogous to a C-scan); described in 7.10.1.1.

- The demonstration that, in the absence of a front wall, multivariate analysis by PCA improves the probability of detection over methods using peak signal values. This may be by an increase in SNR but is always by an improvement in the confusion statistics. (Sections 7.9.1 and 7.9.3). For TFM immersion images the initial measurements are not possible in the presence of a front wall (prior to the application of PCA) and the comparison is not possible. This provides further emphasis of the significance of the previous contribution.

- The development of a robust method using PCP, that demonstrates an almost 100% probability of detection with minimal false alarm rate, to be achievable from multivariate A-scan data; described in section 8.9.3.

- A demonstration of the potential of multivariate analysis for the rapid identification of anomalous regions in masses of regular ultrasonic data resulting from production line tests.

9.2   Further work

The test pieces available for this work are limited and a first suggestion for further work must be a repetition of the techniques on more extensive data sets and ultimately on production data.

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. For TFM images the same assertions cannot be made and this is where the main suggestions for further work are proposed.

Unlike the sectorial data where each image is directly from an angular set of A-scans, the TFM images result from a number of complex mathematical operations constituting the algorithm. In terms of creating an image for analysis this is extremely convenient. For example focusing through the front wall, and weld cap, is accounted for as part of

259

the method. The set of operations do, however, introduce uncertainty. The most obvious evidence of this is the tilt that occurs on some images. This does influence projections with a detrimental effect on image thresholding. Rather than use TFM data the suggestion is, therefore, to use the A-scan data directly from the FMC data set. This becomes equivalent to using sectorial data which is subject to no mathematical operation other than the application of a focal law.

As in the case of TFM image analysis, training sets for each offset now comprise the set of corresponding A-scans. A robust low rank subspace using PCP provides the eigenvectors against which each test observation is projected. Previous work using TFM images demonstrates that front wall effects register as normal background (figures 8.16) the difference now being that artifacts, from TFM, are omitted, resulting in a cleaner image that facilitates thresholding and blob detection.

These conclusions start with a statement that for AUT an advantage of phased arrays is that data is recorded from the entire volume of the weld. It should also be acknowledged that, within the industry, there is considerable experience and expertise in the application of pitch-echo and pitch-catch techniques. It is pointed out that the PCA techniques presented here may also be used with these methods of data capture. In fact the techniques are applicable wherever there is a recording of such periodic data.

9.3 Industrial impact

High profile incidents such as the BP Maconda disaster (2010), the San Bruno pipeline explosion (2005) and the BP Texas City refinery explosion (2005) have placed a stringent focus on regulatory requirements for non-destructive test in general. Stricter regulations are now driving an increase in the recording of all types of non-destructive test data. Using only pipeline manufacture as an example, many infrastructure projects contain vast numbers of welds of identical shape all of which must, for regulatory compliance, safety and insurance be inspected by qualified personnel [142]. Inspection of every image in this way is a time consuming, costly and error-prone operation.

This work has demonstrated the potential of automating the detection of anomalies in weld data. Although the methods are applicable only to welds of a regular structure, such as pipelines, the potential impact of the techniques is not to be underestimated. In particular regular structures are usually the most amenable to automatic data acquisition. Over only a short section of pipeline a phased array, mounted on a suitable robotic manipulator, can

easily generate data for tens of thousands of images a day.

It has been demonstrated that using sectorial data from a contact phased array it is possible to locate anomalies with almost 100% probability of detection. In most cases the suggestion is that false alarms correspond to borderline anomalies that require further investigation. It is not proposed that this method replace the human inspector. A more realistic expectation is that it provides an alternative to be used alongside the qualified inspector. Another potential application applies to the situation where a company contracts out inspection to a third party. Results from automation may now be compared with the reports from the inspection company. Comparison of this data, in terms of confusion statistics, will provide a method of objectively assessing both approaches.

Investigations using TFM images were not as positive as those from a contact phased array. In particular they struggle to produce 90% probability of detection with rather more false alarms. Nevertheless they demonstrate that a weld cap does not present a barrier to producing an underlying image of the weld's structure.

Appendices

Appendix A

Delay and sum beam forming

A.1   Delay and sum beam forming

By applying a set of delays to each element before summing it is possible to enhance the phased array's sensitivity to a wavefront from a specific direction. This delay and sum is a basic operation of the phased array. From figure A.1 it is seen that a wavefront at an angle θ and incident on the reference element (e_0) requires to travels an extra distance $l\sin\theta$ (where x is the pitch of the phased array) to the next element.
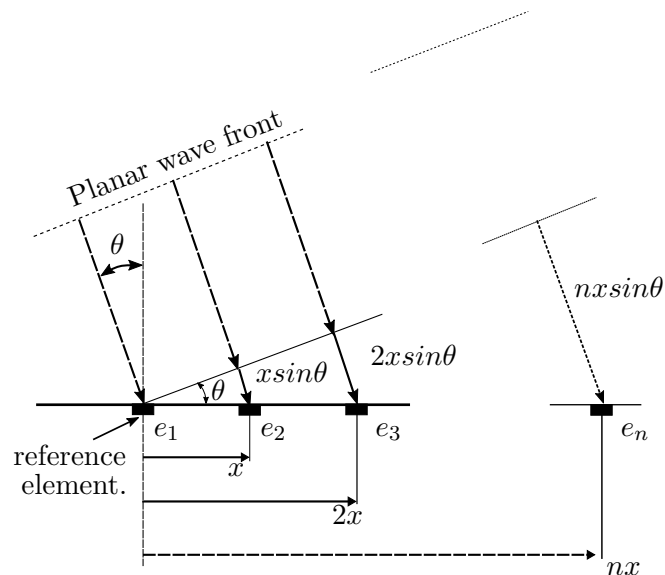

Figure A.1: Delay Calculation

More generally the distance the wave front must travel from the reference element to the next element is:-

$$d = x \sin \theta \tag{A.1}$$

If the speed of sound in the medium is c then the time taken for the wave front to travel this distance is:-

$$t = \frac{x \sin \theta}{c} \tag{A.2}$$

More generally the time taken for the wavefront to travel from the reference element to element m is mt and the output of an N element array for single frequency, ω, and angle, θ is given by:-

$$output = \frac{1}{N} \sum_{m=0}^{N-1} e^{-j\omega mt} \tag{A.3}$$

It is useful to have a method of determining the output for a given L without having to sum each term in the sequence. Substituting k for $-j\omega mt$ and α for e the output sum, S, may be written as:-

$$S = \sum_{k=0}^{N-1} \alpha^k \tag{A.4}$$

Manipulation of the geometric series follows:-

$$S = 1 + \alpha + \alpha^2 + \alpha^3 + ...\alpha^{N-2} + \alpha^{N-1}$$
$$\alpha S = \alpha + \alpha^2 + \alpha^3 + ...\alpha^{N-2} + \alpha^{N-1} + \alpha^N$$
$$S - S\alpha = 1 - \alpha^N$$
$$S = \frac{(1 - \alpha^N)}{(1 - \alpha)} \tag{A.5}$$

Illustrating that equation A.3 is equivalent to:-

$$output = \frac{1}{N} \left(\frac{1 - e^{-j\omega Nt}}{1 - e^{-j\omega t}} \right) \tag{A.6}$$

More generally the frequency response of an N element phased array may be represented

by:-

$$H(j\omega) = \frac{1}{N}\left(\frac{1 - e^{-j\omega Nt}}{1 - e^{-j\omega t}}\right)$$

$$= \frac{1}{N}\left(\frac{e^{-j\omega Nt/2}(e^{j\omega Nt/2} - e^{-j\omega Nt/2})}{e^{-j\omega t/2}(e^{j\omega t/2} - e^{-j\omega t/2})}\right)$$

$$= e^{-j\omega t(N-1)/2}\left(\frac{\sin(\omega Nt/2)}{N\sin(\omega t/2)}\right) \tag{A.7}$$

The magnitude component of A.7 is of the form:-

$$|H(x)| = \frac{1}{M}\frac{\sin(Mx)}{\sin(x)} \tag{A.8}$$

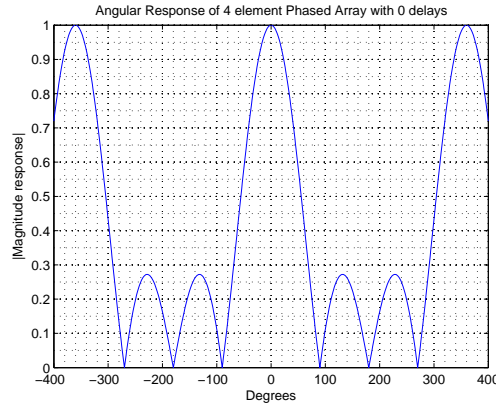An example of this function for a 4 element array ($M = 4$) is illustrated in figure A.2.

Figure A.2: Example response of 4 element array without steering

The magnitude at $x = 0$ is the main lobe. Other maxima at multiples of 2π are grating lobes; these are avoided by restricting x to the range $\pm\pi$. In practice this requires the time delay between a plane wave reaching two adjacent elements of the array to be less than the wavelength (λ) of the sound wave. That is for $d < \lambda$ only the main grating lobe maps to the physical range of $0 \leq 2\theta \leq 2\pi$.

Appendix B

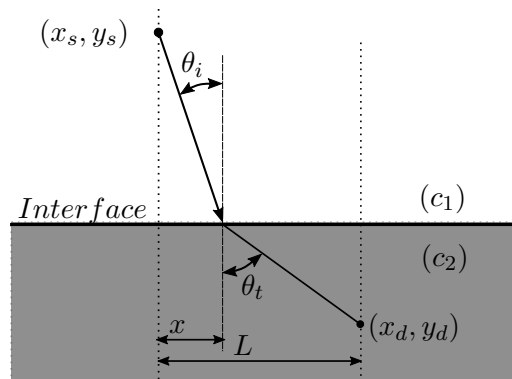# Snell's law derivation

B.1 Snell's law

Figure B.1: Phased array foot print

Figure B.1 provides some elaboration of this situation. Here the point x_s, y_s represents a point source (the centre of the transducer) and x_d, y_d represents a destination point within the body of the test piece. Fermat's principle of least time determines the actual path taken by the sound wave from source to destination. The time taken for a sound field to propagate across a refractive interface between the two points is, in two dimensions:-

$$t = \frac{\sqrt{y_s^2 + x_s^2}}{c_1} + \frac{\sqrt{y_d^2 + (L - x)^2}}{c_2} \tag{B.1}$$

To minimise the time taken it is necessary to set the first derivative of time with respect

to x equal to 0:-

$$\frac{dt}{dx} = \frac{x}{c_1\sqrt{y_s^2 + x_s^2}} + \frac{-(L - x)}{c_2\sqrt{y_d^2 + (L - x)^2}} = 0 \qquad \text{(B.2)}$$

From Figure B.1 it is seen that:-

$$\frac{x}{\sqrt{y_s^2 + x_s^2}} = \sin(\theta_i) \qquad \text{(B.3)}$$

$$\frac{(L - x)}{\sqrt{y_d^2 + (L - x)^2}} = \sin(\theta_r) \qquad \text{(B.4)}$$

so that equation (B.2) becomes:-

$$\frac{dt}{dx} = \frac{\sin(\theta_i)}{c_1} + \frac{\sin(\theta_r)}{c_2} = 0 \qquad \text{(B.5)}$$

B.2   Dual layer - point of interception

Using the arrangement of figure B.2.b with the point x_s, z_s as the reference point, Snell's law may be expressed in its geometric form:-

$$\frac{(x_b - x_s)}{c_1\sqrt{(z_b - z_s)^2 + (x_b - x_s)^2}} = \frac{(x_d - x_b)}{c_2\sqrt{(z_d - z_b)^2 + (x_d - x_b)^2}}$$
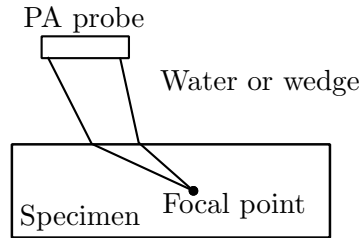
a) Physical set up   b) Refraction at planar interface

Figure B.2: Dual layer set up

Squaring each side gives:-

$$\frac{(x_b - x_s)^2}{(z_b - z_s)^2 + (x_b - x_s)^2} = \frac{\beta(x_d - x_b)^2}{(z_d - z_b)^2 + (x_d - x_b)^2}$$
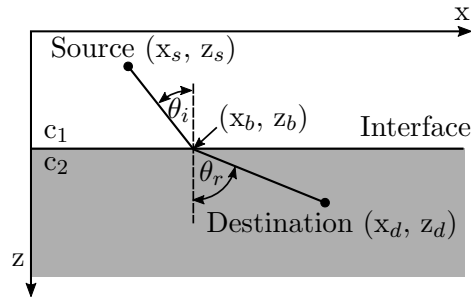
Where:-

$$\beta = \frac{c_1^2}{c_2^2}$$

Re-arrangement leads to the following [143]:-

$$p_4 x_b^4 + p_3 x_b^3 + p_x x_b^2 + p_1 x_b + p_0 = 0 \tag{B.6}$$

Where:-

$$\beta = (c1/c2)^2$$

$$p_4 = 1 - \beta$$

$$p_3 = 2x_d p_4$$

$$p_2 = (z_d - z_b)^2 + x_d^2 - \beta(z_b^2 + x_d^2)$$

$$p_1 = -2x_d z_b^2 \beta$$

$$p_0 = -\beta z_b^2 x_d^2$$

A number of techniques for solving the quadratic polynomial (equation B.2) are available, notably Ferrari's method [41]. However in this work the simplest and most convenient approach is to use the Matlab in-built roots function. Of the four roots x$-b$ is given by the one, and only one, positive real number.

Bibliography

[1] British-Standard-Institution. Quality requirements for fusion welding of metallic materials. Standard BS EN ISO 3834-1:2005, British-Standard-Institution, 2005.

[2] British-Standard-Institution. Methods for ultrasonic examination of welds. Standard BS 3923-1:1986, British-Standard-Institution, 1986.

[3] Ministry of Defence. Def. Stan. 02-773 Minimum Acceptance Standards for Welds in HM Surface Ships and Submarines. Standard, Ministry-of-Defence, 04 2000.

[4] B. Apps, B. Crossland, R. Fenn, and C. Evans. Killer consequencies of defective welds - a plan for prevention. *TWI Bulletin*, Jan 2002.

[5] Mendelu. Non-Destructive Testing (NDT) History of non-destructive testing. `http://web2.mendelu.cz/af_291_projekty2/veso/print.php?page=5524&type=html`, May 2018. Online; Accessed:19-05-2018.

[6] J. Brutz. The evolution of nondestructive testing. *Inspectioneering*, May 2012.

[7] R. DeNale and C.A. Lebowitz. A comparison of ultrasonics and radiography for weld inspection. Technical report, David Taylor Research Center, 1990.

[8] T. Lukomski, T. Stepinski, and J. Kowal. Synthetic aperture focusing technique with virtual transducer for immersion inspection of solid objects. *Insight - Non-Destructive Testing and Condition Monitoring*, 54:623–627, 11 2012.

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

[10] D. Yapp and S.A. Blackman. Recent developments in high productivity pipeline welding. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 26(1), Jan 2004.

[11] E.A. Ginzel. Automated Ultrasonic Testing for pipeline girth welds:a handbook. Technical report, olympus NDT, 2006. Accessed 2015-09-30.

[12] Krautkramer H. Krautkramer J. *Utrasonic Testing of Materials*. New York:Springer Verlag, 1977.

[13] Ginzel E.A. Weld inspection of ultrasonic inspection 2 -Training for nondestructive testing. Technical Report 4, NDTnet, Apr 1998. Online; Accessed:12-06-2015.

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

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

[16] 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.

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

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

[19] 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*, volume 4, pages 2296–3000, 08 2004.

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

[21] S. Bae, L. Udpa, S. Udpa, and T. Taylor. Classification of ultrasonic weld inspection data using principal component analysis. *Review of Progress in Quantitative Nondestructive Evaluation*, 16, 1997.

[22] Simpson G. Bliltz J. *Ultrasonic Methods of Non-destructive Testing*. Chapman and Hall, 1996.

[23] G.K. Ananthasuresh, K.J. Vinoy, S. Gopalakrishnan, K.N. Bhat, and V.K. Aatre. *Micro and Smart Systems*. Wiley India Pvt. Limited, 2010.

[24] C. Guarnaccia, N. Mastorakis, and J. Quartieri. A mathematical approach for wind turbine noise propagation. *International journal of mechanics*, 5:129 – 137, 2011.

[25] Olympus. Technical note. Technical report, Olympus, 2012. Accessed 2016-09-30.

[26] M.V. Brook. *Ultrasonic Inspection Technology Development and Search Unit Design*. John Wiley and Sons, 2012.

[27] ASNT. Nondestructive testing handbook. *American Society for Nondestructive Testing*, 7(3), 2007.

[28] J.A. Brown and G.R. Lockwood. Low-cost, high performance pulse generator for ultrasound imaging. *IEEE Transactions on Ultrasonics, Ferroelectronics and Frequency Control*, 49(6), 2002.

[29] Rabaey J. M. and Chandrakasan A. *Digital Integrated Circuits: A Design Percpective*. Pearson, 2003.

[30] L.E. Kinsler, A.R. Frey, A.B. Coppens, and J.V. Sanders. *Fundamentals of Acoustics*. John Wiley, 2000.

[31] S.C. Wooh and Y. Shi. A simulation study of beam steering characteristics for linear phased arrays. *Journal of Nondestructive evaluation*, 18:39–57, June 1999.

[32] L.W. Schmerr. *Fundamentals of Ultrasonic Phased Arrays*. Springer, 2014.

[33] S. Cohran. Fundamentals of ultrasonic phased arrays. Technical report, researchgate, 2006. Accessed 30-08-2016.

[34] Olympus. NDT-tutorials. Technical report, Olympus, 2012. Accessed 2016-09-30.

[35] S.C. Wooh and Y. Shi. Optimum beam steering of linear phased arrays. *Wave Motion*, 29(3):245 – 265, 1999.

[36] S.C. Wooh and Y. Shi. Influence of phased array element size on beam steering behavior. *Ultrasonics*, 36(6):737 – 749, 1998.

[37] L. Azar, Shi Y., and S.-C. Wooh. Beam focusing behavior of linear phased arrays. *NDT & E International*, 33(3):189 – 198, 2000.

[38] D.H. Turnbull and F.S. Foster. Fabrication and characterisation of transducer elements in two dimensional arrays for medical imaging. *IEEE Trans Ultrasonics, Ferroelectrics, and Frequency Control*, 45(3):667–677, 1998.

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

[40] C. Holmes, B. Drinkwater, and P. Wilcox. Practical experience of phased array technology for power station applications. *Insight - Non-Destructive Testing and Condition Monitoring*, 46:525–528, 9 2004.

[41] D.N. Beshers and A. Oppenheim. The generation of acoustic harmonics by dislocation dipoles and loops. *Journal of Applied Physics*, 52(22):6509–6521, 1981.

[42] Seydel J. Ultrasonic synthetic-aperture focusing techniques. *NDT research techniques in nondestructive testing*, 6, 1982.

[43] S. Bannouf, S. Robert, O. Casula, and C. Prada. Data set reduction for ultrasonic tfm imaging using the effective aperture approach and virtual sources. *Journal of Physics: Conference Series*, 457(1):012007, 2013.

[44] A. Fidahoussen, P. Calmon, M. Lambert, S. Paillard, and S. Chatillon. Imaging of defects in several complex configurations by simulation-helped processing of ultrasonic array data. *Review of Quantitative Nondestructive evaluation*, 20:847–854, 2009.

[45] Deutsch W.A.K., Roye W., Rast H., and Benoist P. High resolution phased array imaging using the total focusing method. 19th *World conference on Non-Destructive Testing*, 2016.

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

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

[48] R. Saranya, J. Daniel, A. Abudhahir, and N. Chermakani. Comparison of segmentation techniques for detection of defects in non-destructive testing images. In *Electronics and Communication Systems (ICECS), 2014 International Conference on*, pages 1–6. IEEE, 2014.

[49] M. Sezgin et al. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1):146–168, 2004.

[50] M. Sezgin and B. Sankur. Selection of thresholding methods for non-destructive testing applications. In *Image Processing, 2001. Proceedings. 2001 International Conference on*, volume 3, pages 764–767. IEEE, 2001.

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

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

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

[54] C.E. Shanon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[55] Y. J. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335 – 1346, 1996.

[56] Hubalek Z. Coefficients of association and similarity, based on binary (presence-absence) data:an evaluation. *Biological Reviews*, pages 669 – 689, 11 1982.

[57] F.J. 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.

[58] F.J. Provost, T. Fawcett, and R. Kohavi. The case against accuracy estimation for comparing induction algorithms. In *ICML*, volume 98, pages 445–453, 1998.

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

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

[61] K. Zuva and T. Zuva. Evaluation of information retrieval systems. *International journal of computer science & information technology*, 4(3):35 – 43, June 2012.

[62] D. A. Jeni, J. F. Cohn, and F. De La Torre. Facing imbalanced data recommendations for the use of performance metrics. In *International Conference on Affective Computing and Intelligent Interaction Workshops*, volume 15, pages 245–251, 2013.

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

[64] Han K.T.M. and Uyyanonvara B. A survey of blob detection algorithms for biomedical images. In *2016 7th International conference of Information Communication Technology*, pages 57–60, 2016.

[65] M. Welvaert and Y. Rosseel. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fmri data. *PLOS ONE*, 11 2013.

[66] E.L. Villaverde, S. Robert, and Prada 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.

[67] A.C. Rencher and W.F. Christensen. *Methods of multivariate analysis*. Wiley, 2012.

[68] F. Lingvall and T. Stepinski. Ultrasonic characterization of defects. Technical report, Uppsala University, 02 1999.

[69] R. Gnanadesikan. *Methods for Statistical Data Analysis of Multivariate Observations*. Wiley, 1997.

[70] E. Anderson. The irises of the Gaspé peninsula. *Bulletin of the American Iris Society*, 59, 1935.

[71] T.F. Karim, M.S.H. Lipu, L. Rahman, and F. Sultana. Face recognition using pca-based method. In *2010 IEEE International Conference on Advanced Management Science(ICAMS 2010)*, volume 3, pages 158–162, July 2010.

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

[73] M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

[74] M.A. Turk and A.P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.

[75] B. Cattell. The scree test for the number of factors. *Multiv Behav Res*, 1:245–276, 1966.

[76] A. Bhattacharjee, W.G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, et al. Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences*, 98(24):13790–13795, 2001.

[77] K. Yata and M. Aoshima. Effective pca for high-dimension, low-sample-size data with singular value decomposition of cross data matrix. *Journal of Multivariate Analysis*, 101(9):2060 – 2077, 2010.

[78] S. Jung and J.S. Marron. PCA consistency in high dimension, low sample size context. *The Annals of Statistics*, 37(6B):4104–4130, 12 2009.

[79] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58, July 2009.

[80] J. Hardin and D.M. Rocke. Outlier detection in the multiple cluster setting using the minimum covariance determinant estimator. *Computational Statistics and Data Analysis*, 44(4):625 – 638, 2004.

[81] Lewis T. Barnett V. *Outliers in statistical data*. John Wiley, 1994.

[82] M.L. Shyu, S.C. Chen, K. Sarrinnapakorn, and L.W. Chang. A novel anomaly detection scheme based on principal component classifier. Technical report, DTIC Document, 2003.

[83] F. Angiulli, F. Basta, and C. Pizzuti. Distance based detection and prediction of outliers. *IEEE Transactions on Knowledge and Data Engineering*, 18(2), 2006.

[84] E.M. Knox and R.T. Ng. Algorithms for mining data-distance based outliers in large datasets. In *Proceedings of the 24th VLDB Conference*, 1998.

[85] P.R. Anand and T.K. Kumar. PCA based anomaly detection. *International Journal of research in Advent Technology*, 2(2), 2014.

[86] M.M. Breunig, H-P. Kriegel, R.T. Ng, and J. Sander. Lof: Identifying density-based local outliers. *SIGMOD Rec.*, 29(2):93–104, May 2000.

[87] G. Snedecor, G.W. Williams. *Statistical Methods*. Iowa State University Press, 1989.

[88] P.C Mahalanobis. On the generalised distance in statistics. In *Proceedings of the National Institute of Science of India*, volume 2, pages 49–55, 1936.

[89] B.F.J. Manly. *Multivariate statistical methods, a primer*. Chapman Hall, 2004.

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

[91] Wichern D.W. Johnson R.A. *Applied Multivariate Statistical Analysis*. Prentice Hall, 1998.

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

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

[94] R. Ledesma and P. Valero-Mora. Determining the number of factors to retain in efa. *Practical Assessment, Research and Evaluation*, 12(2), 02 2007.

[95] H.F. Kaiser. The application of electronic computers to factor analysis. *Ed. Psychol. Meas.*, 20:141–151, 1960.

[96] E. Acuna and C.A. Rodriguez. Meta analysis study of outlier detection methods in classification. In *Proceedings IPSI*, 2004.

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

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

[99] P.J. Rousseeuw. Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, B:283–297, 1985.

[100] P.J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Society*, 79:871–880, 1985.

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

[102] N. Badr and N.A. Noureldien. Examining outlier detection performance for principal components analysis method and its robustification methods. *International Journal of Advances in Engineering and Technology*, 6(2):573–582, May 2013.

[103] D. A. Jackson and Chen Yong. Robust principal component analysis and outlier detection with ecological data. *Environmetrics*, 15:129–139, 2004.

[104] A. S. Kosinski. A procedure for the detection of multivariable outliers. *Computational Statistics and Data Analysis*, 29:145–161, 1998.

[105] A.I. Marden. Some robust estimates of principal components. *Statistics and Probability Letters*, 43:349–359, 1999.

[106] P.J. Rousseeuw. Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, B:283–297, 1985.

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

[108] Y.J. Lee, Y.R. Yeh, and Y.C.F. Wang. Anomaly detection via online oversampling principal component analysis. *IEEE Transactions on Knowledge and Data Engineering*, 25(7):1460–1470, July 2013.

[109] R. Kwitt and U. Hofmann. Unsupervised anomaly detection in network traffic by means of robust pca. In *Computing in the Global Information Technology, 2007. ICCGI 2007. International Multi-Conference on*, pages 37–37, March 2007.

[110] A. Drobchenko, J-K. Kämäräinen, L. Lensu, J. Vartiainen, and H. Kälviäinen. Thresholding-based detection of fine and sparse details. *Frontiers of Electrical and Electronic Engineering*, 6(2):328, 2011.

[111] P.L. Rosin. Unimodal thresholding. *Pattern Recognition*, 34(11):2083 – 2096, 2001.

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

[113] 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/.

[114] K. Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science*, 6(2):559–572, 1901.

[115] H. Hotelling. Analysis of a complex statistical variable into components. *Journal of Educational Psychology*, 24:417–441, 1933.

[116] K.I. Diamantaras. *Principal Component Neural Networks (Theory and applications)*. John Wiley and Sons, 1996.

[117] C. Eckardt and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

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

[119] P.J. Huber. *Robust Statistics*. Wiley:New York, 1981.

[120] E. Cermeño, A. Pérez, and J.A. Sigüenza. Intelligent video surveillance beyond robust background modelling. *Expert Systems with Applications*, 91:138–149, January 2018.

[121] N.A. Cambell. Robust procedures in multivariate analysis I:robust covariance estimation. *Applied Statistics*, 3(29):231–237, 1980.

[122] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

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

[124] A. Baccini, P. Besse, and A. Falguerolles. The l1-norm PCA and a heuristic approach. *Ordinal and Symbolic Data Analysis*, pages 359–368, 1996.

[125] Q. Ke and T. Kanade. Robust l1 norm factorization in the prescence of outliers and missing data by alternative complex programming. *Proceedings IEEE Conf. Computer Vision and Convex Programming*, June 2003.

[126] Q. Ke and T. Kanade. Robust subspace computation using l1-norm. *Technical Report CMU-CS-03-172*, 2003.

[127] C. Ding, D. Zhou, He X., and Zha H. R1-pca: Rotational invariant l1-norm Principal Component Analysis for robust subspace factorization. *Proc. 23rd Int'l Conf. Machine Learning*, 2006.

[128] G. Golub and C.V. Loan. *Matrix Computation*. John Hopkins Univ. Press, 1996.

[129] Y.W. Park and D. Klabjan. Iteratively reweighted least squares algorithms for l1-norm principal component analysis. *arXiv e-prints*, Sep 2016. Available at https://arxiv.org/abs/1609.02997v2.

[130] 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*, 59:797 – 829, 06 2006.

[131] Venkat C., Sujay S., Pablo A.P., and Alan S.W. Rank-sparsity incoherence for matrix decomposition. *SIAM Journal on Optimization*, 21(2):572–596, 2011.

[132] J. Wright, Y. Peng, A. Ganesh, and S. Rao. Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization. *Neural Information Processing Systems*, 2009.

[133] Becker S., Candès E., and Grant M. Templates for convex cone problems with applications to sparse signal recovery, 2010. Available at http://github.com/cvxr/TFOCS/releases.

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

[135] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011.

[136] B. Peleato and I. Papusha. Robust principal component analysis (and applications), 2014. Available at https://statweb.stanford.edu/ candes/math301/Lectures/rpca.pdf.

[137] N. Oliver and A. Pentland. A bayesian computer vision system for modelling human interactions. *International Conference on Computer Vision systems*, 1999.

[138] Aybat N.S., Goldfarb D., and Ma S. Efficient algorithms for robust and stable principal component pursuit problems. *Comput Optim Appl*, 58(1), May 2014.

[139] I.L. Hyeung and JungWoo L. Online update techniques for projection based robust principal component analysis. *ICT Express*, 1(2):59 – 62, 2015.

[140] B. Cyganek and M. Woźniak. Efficient real-time background detection based on the PCA subspace decomposition. In L. Rutkowski, M. Korytkowski, R. Scherer,

R. Tadeusiewicz, L.A. Zadeh, and J.M. Zurada, editors, *Artificial Intelligence and Soft Computing*, pages 485–496, Cham, 2017. Springer International Publishing.

[141] Fan Jiangqing and Wang Weichen. Asymptotics of empirical eigen-structure for ultra-high dimensional spikes covariance model. *arXiv e-prints*, Feb 2015. Available at https://arxiv.org/abs/1502.04733v2/.

[142] Black Arch Partners. An overview of nondestructive testing white paper. `www.blackarchpartners.com/industry-approach/energy`.

[143] S. Ho-Chul, R. Prager, H. Gomersall, N. Kingsbury, G. Treece, and A. Gee. Estimation of speed of sound in dual-layered media using medical ultrasound image deconvolution. *Ultrasonics*, 50(7):716 – 725, 2010.